Nonlinear regression model checking via local polynomial smoothing

Ricardo Cao and Salvador Naya (**email:** salva@udc.es)

Department of Mathematics, University of Coruña. Spain.

Address: Escola Politecnica Superior, Mendizabal, 15403 Ferrol. Spain.

January 30, 2007

Abstract

A goodness-of-fit test statistic for nonlinear regression models based on local polynomial estimation is proposed in this paper. The criterion used to construct the test is the distance between the parametric fit and the nonparametric regression estimation. The good performance of the test is shown via a simulation study. The method is applied to check a logistic mixture regression model for real data coming from a thermal analysis problem.

Key words: Goodness-of-fit, hypothesis testing, logistic mixture model, nonparametric regression

Acknowledgements: Research supported by the MEC Grant MTM2005-00429 (ERDF included).

Running head: Regression checking via local polynomials.

1 Introduction

A relevant statistical problem in real data applications is to estimate the regression function of a response variable given some covariates. Although linear models are spreadly used in many practical setups these models are often not flexible enough to capture the complex information contained in the data. For this reason, plenty of other parametric models can be used to explain nonlinear relations. A previous step before fitting such nonlinear models to the data is to perform goodness-of-fit tests for the proposed model.

There are two different approaches in the literature for model checking in a regression context. The first one consists of using nonparametric smoothing techniques for estimating the regression function and compare this estimation with that coming from a parametric fit. Some discrepancy measure between both estimators has to be introduced and, as a consequence, a test statistic is defined. Some relevant references in this context are HÄRDLE and MAMMEN (1993), GONZÁLEZ-MANTEIGA and CAO (1993), STUTE and GONZÁLEZ-MANTEIGA (1996), GONZÁLEZ-MANTEIGA and VILAR (1995), ZHENG (1996), HÄRDLE, MAMMEN and MÜLLER (1998), LI and WANG (1998), RODRÍGUEZ-CAMPOS, GONZÁLEZ-MANTEIGA and CAO (1998), DETTE and MUNK (1998), ALCALÁ, CRISTÓBAL and GONZÁLEZ-MANTEIGA (1999), HÄRDLE and KNEIP (1999), VILAR and GONZÁLEZ-MANTEIGA (2000) and GONZÁLEZ-MANTEIGA and PÉREZ-GONZÁLEZ (2006).

The second approach is based on empirical regression process theory and uses well known functionals (as the Kolmogorov-Smirnov or the Cramer-von Mises) of this process to define goodness-of-fit test statistics. One of the first references in this setup is the paper by STUTE (1997). A sample of other relevant references is STUTE, GONZÁLEZ-MANTEIGA and PRESEDO-QUINDIMIL (1998), STUTE, THIES and ZHU (1998), KAUERMANN and TUTZ (2001), DIEBOLT and ZUBER (2001), STUTE and ZHU (2002), LIN, WEI and YING (2002), NEUMEYER and DETTE (2003), ZHU (2003), KHMALADZE and KOUL (2004), STUTE and ZHU (2005) and ESCANCIANO (2006).

The smoothing approach for nonlinear model checking is adopted in this paper. A test statistic based on the L_2 distance between a local linear estimator and a parametric fit is introduced in Section 2. The limit distribution of the test is presented in Section 3. The performance of the test is examined in Section 4 by means of a simulation study. Section 5 is devoted to a real data application in the context of thermal analysis. Finally,

the proofs of the asymptotic results are collected in Section 6.

2 Goodness-of-fit test

We concentrate ourselves in the fixed design regression context:

$$Y_i = m(t_i) + \varepsilon_i, \quad i = 1, 2, ..., n, \quad \text{with } E(\varepsilon_i) = 0,$$

where m is the regression function of Y given T, ε_i are zero mean, independent and identically distributed random variables and the design satisfies $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$.

We consider a (possibly nonlinear) parametric model for m, namely $\mathcal{M} = \{m_\theta(\bullet)/\theta \in \Theta\}$, where Θ is a subset of $\mathbb{R}^k$. The hypothesis testing under study is $H_0 : m \in \mathcal{M}$ (i.e. $\exists \theta_0 \in \Theta / m = m_{\theta_0}$) versus the alternative $H_1 : m \notin \mathcal{M}$.

Starting from the sample $\{(t_j, Y_j)\}_{j=1}^n$, we consider, under H_0, the least squares estimator of θ_0. This is the $k-$dimensional vector, $\widehat{\theta}$, that minimizes, in θ, the sum of squared residuals:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \psi_n(\theta), \tag{1}$$

with

$$\psi_n(\theta) = \sum_{j=1}^{n} \left(Y_j - m_\theta\left(t_j\right)\right)^2. \tag{2}$$

The goodness-of-fit test is defined by some discrepancy measure between the parametric regression estimator, $m_{\widehat{\theta}}$, and some nonparametric estimator, $\widehat{m}_h$. A typical choice for this discrepancy is $D = d^2(\widehat{m}_h, m_{\widehat{\theta}})$, where d is a suitable functional distance. Along this paper d will be the L_2 distance and the nonparametric estimator, $\widehat{m}_h$, will be the local polynomial estimator proposed by FAN and GIJBELS (1996):

$$\widehat{m}_h\left(t\right) = \sum_{i=1}^{n} W_0^n\left(\frac{t - t_i}{h}\right) Y_i \tag{3}$$

where h is the smoothing parameter,

$$W_0^n(u) = e_1^T S_n^{-1} \left(1, hu, \ldots, (hu)^p\right)^T K(u)/h, \tag{4}$$

K is the kernel function and

$$\mathbf{S}_n = \left(s_{j,l}^{(n)}\right)_{0\leq j,l\leq p},$$

with

$$s_{j,l}^{(n)} = S_{n,j+l},$$

$$S_{n,j} = \sum_{i=1}^{n} K_h(t-t_i)(t_i-t)^j. \tag{5}$$

Thus, the final expression for the test statistic is

$$D = d^2(\widehat{m}_h, m_{\widehat{\theta}}) = \sum_{j=1}^{n} \left(\widehat{m}_h\left(t_j\right) - m_{\widehat{\theta}}\left(t_j\right)\right)^2. \tag{6}$$

3 Asymptotic properties

The following conditions will be needed to obtain the asymptotic normality of the parametric estimator and the limit distribution of the test statistic under the null hypothesis

A1. $nh^2 \to \infty$ and $nh^{7/2} \to 0$

A2. K is a symmetric continuously differentiable density function with support on $[-1,1]$.

A3. $E\left(\varepsilon^4\right) < \infty$, where $\sigma^2 = E\left(\varepsilon^2\right)$.

A4. The design is asymptotically equispaced in $[0,1]$, i.e. $\max_{2\leq j\leq n} \left|t_j - t_{j-1} - \frac{1}{n}\right| = o\left(\frac{1}{n}\right)$.

A5. $\sup_{t\in[0,1]} \left|m_{\theta_1}(t) - m_{\theta_2}(t)\right| \leq C_1 \left\|\theta_1 - \theta_2\right\|$.

A6. The matrix $\mathbf{H}_n = \mathbf{H}_n\psi_n\left(\theta_0, \vec{t}, \vec{Y}\right) = \left(h_{lr}^{(n)}\right)$, with $h_{lr}^{(n)} = \left.\frac{\partial^2\psi_n}{\partial\theta_{il}\partial\theta_r}\right|_{\theta=\theta_0}$, is nonsingular.

A7. The function $m_\theta(t)$ is three times continuously differentiable in θ, for every $t \in [0,1]$, with bounded third partial derivatives in θ.

A.8 The function $\frac{\partial}{\partial\theta_i} m_\theta(t)$ is uniformly continuous in t, for every $i = 1, 2, \ldots, k$.

The asymptotic limit distribution of the least squares estimator is now stated.

Theorem 1 *Let us consider the estimator $\widehat{\theta}$ defined in (1) and assume Conditions A1-A8. Then, if H_0 holds,*

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} \mathbf{N}_k(0, \mathbf{S}), \tag{7}$$

where $S = \mathbf{H}^{-1}\Sigma\left(\mathbf{H}^{-1}\right)^T$, $\mathbf{H} = (h_{lr})$, *with*

$$h_{lr} = \begin{cases} 2\int_0^1 \left(\frac{\partial}{\partial\theta_l} m_\theta(t)\right)^2 dt \text{ if } l = r \\ 2\int_0^1 \frac{\partial}{\partial\theta_l} m_\theta(t) \frac{\partial}{\partial\theta_r} m_\theta(t) dt \text{ if } l \neq r \end{cases} \tag{8}$$

and $\Sigma = (\sigma_{lr})_{1\leq l,r\leq k}$, *with*

$$\sigma_{lr} = \sigma^2 \int \left.\frac{\partial}{\partial\theta_l} m_\theta(t)\right|_{\theta=\theta_0} \left.\frac{\partial}{\partial\theta_r} m_\theta(t)\right|_{\theta=\theta_0} dt.$$

Now, the limit distribution of the test statistic is also established.

Theorem 2 *Let us consider the local polynomial estimator defined in (3) and the test statistic given in (6) and assume Conditions A1-A8. Under* H_0,

$$\sqrt{n^2 h}\left(D - \frac{\sigma^2}{nh}\int_{-1}^{1} K^2\right) \xrightarrow{d} N\left(0, 2\sigma^4 \int_{-1}^{1} (K * K)^2\right), \tag{9}$$

where $*$ *denotes convolution.*

4 Simulation study

The procedure presented above has been applied to test a logistic mixture model with a known number, d, of components:

$$m_\theta(t) = \sum_{i=1}^{d} w_i g(a_i + b_i t),$$

where

$$g(x) = \frac{e^x}{1+e^x}$$

and the parameter vector is $\theta = (w_1, a_1, b_1, \ldots, w_d, a_d, b_d)^T \in \Theta = (\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R})^d$.

For the null hypothesis case we have set $d = 2$ and the value of the parameter vector was $\theta_0 = (5, 12, -4, 4, 14, -2)$. The function $m_0(t) = m_{\theta_0}(t)$ has been plotted in Figure 1.

A total number of n ($n = 400, 500, 600$) equispaced data, t_i ($i = 1, 2, \ldots, n$) have been considered in the interval $[0, 12]$. The response variable has been generated according to $Y_i = m(x_i) + \varepsilon_i$ (with $m = m_0$), where ε_i are iid $N(0, \sigma^2)$ random variables with $\sigma = 0.1$. The number of trials in the simulation was 1000.

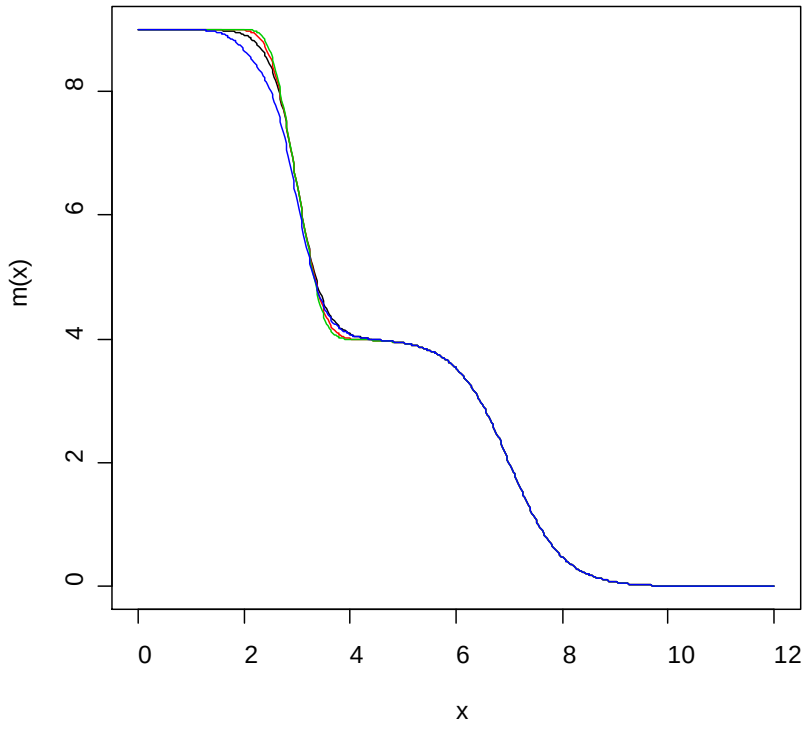

Figure 1: Regression functions m_0 (black curve), m_1 (red curve), m_2 (green curve) and m_3 (blue curve) for the null hypothesis case.

The following alternative hypothesis have been considered:

$$\begin{aligned} m_1 &= 5f\left((12-4x)(0.5(x-3)^2+1)\right)+4f(14-2x), \\ m_2 &= 5f\left((12-4x)((x-3)^2+1)\right)+4f(14-2x), \\ m_3 &= 4.5f(12-4x)+4f(14-2x)+0.5f(10-5x). \end{aligned}$$

The function m_1 is very close to m_0. It slightly differs in a small perturbation in the first logistic component. The second alternative model, m_2, incorporates a somewhat larger perturbation in the same component. Finally, the function m_3 includes and additional logistic component, which reflects common alternative models, in practice, as those studied in Section 5.

The normal approximation for the test statistic null distribution has been used:

$$D \overset{d}{\simeq} N\left(\frac{\sigma^2}{nh}\int_{-1}^{1}K^2, \sqrt{\frac{2\sigma^4\int_{-1}^{1}(K*K)^2}{n^2h}}\right),$$

A local linear estimator (i. e., $p = 1$) was used for the nonparametric regression smoother with several subjective bandwidths: $h = 0.11$, 0.13, 0.15. The corresponding bandwidths were also used to compute the nonparametric residuals $\hat{\varepsilon}_i = Y_i - \hat{m}_h(t_i)$. The sample variance of these residuals, $\hat{\sigma}^2$, was used to estimate the unknown term σ^2 in the expression for D. The Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)^+$ was used.

Tables 1 and 2 collect the acceptance proportions for the null hypothesis and the three alternative models using the three bandwidths mentioned above, sample size $n = 500$ and significance levels $\alpha = 0.01$, 0.05. It is clearly seen that the test is conservative but it has a large power, especially for alternative models m_2 and m_3. It is also evident that the smoothing parameter has an important role, especially when examining the power of the test.

	$h = 0.11$	$h = 0.13$	$h = 0.15$
m_0	1	1	1
m_1	0.988	0.983	0.854
m_2	0.886	0.737	0.251
m_3	0.729	0.640	0.401

Table 1. Acceptance proportions for the four models, using bandwidths $h = 0.11$, 0.13, 0.15, sample size $n = 500$ and significance level $\alpha = 0.01$.

	$h = 0.11$	$h = 0.13$	$h = 0.15$
m_0	0.995	0.999	0.959
m_1	0.502	0.593	0.071
m_2	0.275	0.077	0.017
m_3	0.176	0.084	0.033

Table 2. Acceptance proportions for the four models, using bandwidths $h = 0.11$, 0.13, 0.15, sample size $n = 500$ and significance level $\alpha = 0.05$.

To evaluate the influence of the sample size a new batch of simulations have been run for $n = 400$, 500, 600. The acceptance rates, corresponding to $\alpha = 0.01$, for models m_0, m_1, m_2 and m_3 are collected in Tables 3-6. The figures in the tables show that the test slightly reduces its conservativeness as the sample size grows. The power of the test

increases significantly with the sample size. The choice of the smoothing parameter is clearly an important issue.

	$n = 400$	$n = 500$	$n = 600$
$h = 0.11$	1	1	1
$h = 0.13$	1	1	0.999
$h = 0.15$	1	1	0.969

Table 3. Acceptance proportions for m_0, using bandwidths $h = 0.11, 0.13, 0.15$, sample size $n = 400, 500, 600$ and significance level $\alpha = 0.01$.

	$n = 400$	$n = 500$	$n = 600$
$h = 0.11$	1	0.988	0.866
$h = 0.13$	1	0.983	0.594
$h = 0.15$	1	0.854	0.039

Table 4. Acceptance proportions for m_1, using bandwidths $h = 0.11, 0.13, 0.15$, sample size $n = 400, 500, 600$ and significance level $\alpha = 0.01$.

	$n = 400$	$n = 500$	$n = 600$
$h = 0.11$	1	0.886	0.222
$h = 0.13$	0.998	0.737	0.047
$h = 0.15$	0.993	0.251	0.002

Table 5. Acceptance proportions for m_2, using bandwidths $h = 0.11, 0.13, 0.15$, sample size $n = 400, 500, 600$ and significance level $\alpha = 0.01$.

	$n = 400$	$n = 500$	$n = 600$
$h = 0.11$	0.998	0.729	0.142
$h = 0.13$	0.997	0.640	0.073
$h = 0.15$	0.982	0.401	0.006

Table 6. Acceptance proportions for m_3, using bandwidths $h = 0.11, 0.13, 0.15$, sample size $n = 400, 500, 600$ and significance level $\alpha = 0.01$.

5 Thermal analysis application

We now apply the logistic mixture model studied in the previous section to a data set from a thermal analysis context. A TGA curve (thermogravimetric analysis) was obtained from a polyurethane sample using a heating ramp of 10 ^{o}C/min and a purge of 50 mL/min of argon. The TGA curve, shown in Figure 2, correspond to the polyurethane mass along time in the thermogravimetric experiment. For more details about the logistic mixture model in thermal analysis see NAYA (2003), NAYA, CAO and ARTIAGA (2003), CAO, NAYA, ARTIAGA, GARCIA and VARELA (2004) and NAYA, CAO, LÓPEZ-DE-ULLIBARRI, ARTIAGA, BARBADILLO and GARCIA (2007).

A sample of $n = 273$ equispaced points in the time domain $[0, 20000]$, in seconds, was available. The bandwidth has been chosen using the plug-in method proposed by RUPPERT, SHEATHER and WAND (1995). The value obtained was $h = 127$. The test statistic has been computed, using the same prescriptions as in the previous section, for testing the null hypothesis that the data come from a four-component logistic mixture model. The normal approximation gives a p-value for the test of $p = 0.6337$, which lead to acceptance of a four component logistic mixture model. A three-component logistic mixture model has also been tested, finding an approximate p-value of $p = 0$, which implies rejection of the reduced model.

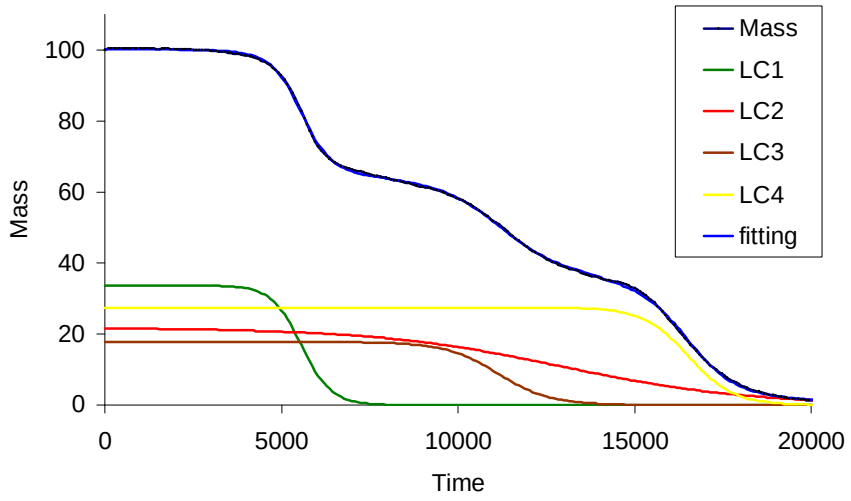

Figure 2: TGA output (black curve), logistic mixture model (blue curve) and logistic components (green, yellow, red and brown curves) for the polyurethane data.

The estimated values of the parameters for the four-component logistic model are collected in Table 7, while the parametric fit and the four logistic components are also plotted in Figure 2. The logistic mixture fit for this data has a nice interpretation in terms of the thermal behaviour of the material. The steps depicted by the four components correspond to different polymer degradation processes. The middle points of these steps are located at the values -$\frac{\hat{a}_i}{\hat{b}_i}$, i. e., 5559, 12907, 11152, 16511.

	$\hat{w}_i$	$\hat{a}_i$	$\hat{b}_i$
$i=1$	33.6487	13.1199	−0.00236
$i=2$	21.6443	4.9047	−0.00038
$i=3$	17.3087	15.1678	−0.00136
$i=4$	27.3987	26.0876	−0.00158

Table 7. Estimated parameters of the four-component logistic mixture model for the polyurethane data.

6 Proofs

The proofs of the two asymptotic results presented in the previous section are given now.

6.1 Proof of Theorem 1

Recall (2) and the definition of H_n in Condition A6. Since $\widehat{\theta}$ is the minimizer of ψ_n, Condition A7 implies that a second order Taylor expansion can be obtained:

$$\overrightarrow{0} = \nabla\psi_n\left(\widehat{\theta}\right) = \nabla\psi_n\left(\theta_0\right) + \mathbf{H}_n(\theta_0)\left(\widehat{\theta} - \theta_0\right) + R_n, \tag{10}$$

with

$$\nabla\psi_n\left(\theta\right) = \left(\frac{\partial\psi_n\left(\theta\right)}{\partial\theta_i}\right)_{1\leq i\leq k}^{t}$$

and

$$R_n = \left(R_{n,1}, R_{n,2}, \ldots, R_{n,k}\right)^T,$$

with

$$R_{n,i} = \frac{1}{2}\sum_{j,l=1}^{k}\frac{\partial^3\psi_n}{\partial\theta_i\partial\theta_j\partial\theta_l}\left(\widetilde{\theta}\right)\left(\widehat{\theta}_j - \theta_{0,j}\right)\left(\widehat{\theta}_l - \theta_{0,l}\right),\ 1\leq i\leq k,$$

for some $\widetilde{\theta}$ between $\widehat{\theta}$ and θ_0.

Now using Condition A6 in (10) gives

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) = A + B \tag{11}$$

where

$$A = -\left(\frac{1}{n}\mathbf{H}_n\right)^{-1}\frac{1}{\sqrt{n}}\nabla\psi_n(\theta_0), \tag{12}$$

$$B = -\left(\frac{1}{n}\mathbf{H}_n\right)^{-1}\frac{1}{\sqrt{n}}R_n(\theta). \tag{13}$$

The proof will conclude by showing that $A \xrightarrow{d} N(0,\mathbf{S})$ and $B = o_P\left(\sqrt{n}\left\|\widehat{\theta}-\theta_0\right\|_2\right)$.

First of all, Cramér-Wold device is a useful tool to prove convergence of distribution of the term

$$\frac{1}{\sqrt{n}}\nabla\psi_n(\theta_0) = -2\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(Y_j - m_{\theta_0}(t_j)\right)\begin{pmatrix}\frac{\partial}{\partial\theta_1}m_\theta(t_j)\\ \vdots \\ \frac{\partial}{\partial\theta_{3k}}m_\theta(t_j)\end{pmatrix}_{\theta=\theta_0}.$$

For any fixed vector $a \in \mathbb{R}^k$, the conditions in Lyapunov Central Limit Theorem (see PETROV (1995), p.126) are easily checked and thus, it can be easily proved that

$$a^T\xi_n \xrightarrow{d} \mathbf{N}_k\left(\overrightarrow{0}, a^t\Sigma a\right).$$

As a consequence,

$$\frac{1}{\sqrt{n}}\nabla\psi_n\left(\theta_0\right) \xrightarrow{d} \mathbf{N}_k\left(\overrightarrow{0},\Sigma\right). \tag{14}$$

The elements $h_{lr}^{(n)}$ of the Hessian matrix $\mathbf{H}_n$ are given by:

$$h_{ll}^{(n)} = 2\sum_{j=1}^{n}\left[\left(\frac{\partial}{\partial\theta_l}m_\theta(t_j)\right)^2 - \left(Y_j - m_{\theta_0}(t_j)\right)\frac{\partial^2}{\partial\theta_l^2}m_\theta(t_j)\right]\Bigg|_{\theta=\theta_0} \tag{15}$$

$$h_{lr}^{(n)} = 2\sum_{j=1}^{n}\left[\frac{\partial}{\partial\theta_l}m_\theta(t_j)\frac{\partial}{\partial\theta_r}m_\theta(t_j) - \left(Y_j - m_{\theta_0}(t_j)\right)\frac{\partial^2}{\partial\theta_l\partial\theta_r}m_\theta(t_j)\right]\Bigg|_{\theta=\theta_0} \tag{16}$$

Using Conditions A4 and A8, it is easy to prove

$$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta_l}m_\theta(t_j)\right)^2 \underset{n\to\infty}{\to} \int_0^1\left(\frac{\partial}{\partial\theta_l}m_\theta(t)\right)^2 dt \tag{17}$$

and

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_l}m_\theta(t_j)\frac{\partial}{\partial\theta_r}m_\theta(t_j) \underset{n\to\infty}{\to} \int_0^1 \frac{\partial}{\partial\theta_l}m_\theta(t)\frac{\partial}{\partial\theta_r}m_\theta(t)dt \tag{18}$$

On the other hand, standard variance calculations and Tchebychev inequality imply

$$\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \left.\frac{\partial^2}{\partial\theta_l^2}m_\theta(t_j)\right|_{\theta=\theta_0} \xrightarrow{P} 0, \tag{19}$$

$$\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j \left.\frac{\partial^2}{\partial\theta_l\partial\theta_r}m_\theta(t_j)\right|_{\theta=\theta_0} \xrightarrow{P} 0. \tag{20}$$

Finally using (17), (18), (19) and (20) in (15) and (16), it follows that

$$\frac{1}{n}\mathbf{H}_n \xrightarrow{P} \mathbf{H}. \tag{21}$$

Now, using (14) and (21) the limit distribution of (12) is easily derived:

$$A \xrightarrow{d} \mathbf{N}_k(0,\mathbf{S}). \tag{22}$$

To deal with the second term in (11), we first bound each component of the vector R_n using Cauchy-Schwarz inequality:

$$|R_{n,i}| \leq \frac{1}{2}\left[\sum_{j,l=1}^{k}\left(\frac{\partial^3\psi_n}{\partial\theta_i\partial\theta_j\partial\theta_l}\left(\widetilde{\theta}\right)\right)^2\right]^{1/2}\left\|\widehat{\theta}-\theta_0\right\|_2^2 \tag{23}$$

As a consequence, using Condition A7, (21) and (23), the term in (13) is

$$B = O_P\left(\frac{1}{\sqrt{n}}\left\|\widehat{\theta}-\theta_0\right\|_2^2\right) = o_P\left(\sqrt{n}\left\|\widehat{\theta}-\theta_0\right\|_2\right)$$

and the proof is finished.

6.2 Proof of Theorem 2

We follow the steps of the proof of Theorem 2.1 in GONZÁLEZ-MANTEIGA and CAO (1993). Starting from (6), standard algebra gives

$$D = d^2\left(\widehat{m}_h, m_{\widehat{\theta}}\right) = A + B + C,$$

where

$$\begin{aligned}
A &= \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{m}_h\left(t_j\right) - m_{\theta_0}\left(t_j\right)\right)^2, \\
B &= \frac{1}{n}\sum_{j=1}^{n}\left(m_{\theta_0}\left(t_j\right) - m_{\widehat{\theta}}\left(t_j\right)\right)^2, \\
C &= \frac{2}{n}\sum_{j=1}^{n}\left(\widehat{m}_h\left(t_j\right) - m_{\theta_0}\left(t_j\right)\right)\left(m_{\theta_0}\left(t_j\right) - m_{\widehat{\theta}}\left(t_j\right)\right).
\end{aligned}$$

Condition A5 and Theorem 1 imply that

$$B = O_p\left(\frac{1}{n}\right). \tag{24}$$

On the other hand, Cauchy-Schwarz inequality imply

$$|C| \leq 2\sqrt{AB}. \tag{25}$$

Consequently, the only dominant term for the asymptotic null distribution of the test statistic is A and the only fact that remains to prove is

$$\sqrt{n^2 h}\left(A - \frac{\sigma^2}{nh}\int_{-1}^{1} K^2\right) \xrightarrow{d} N\left(0, 2\sigma^4 \int_{-1}^{1} (K * K)^2\right). \tag{26}$$

Recall expression (3). Under H_0, the term A can be decomposed as follows:

$$\begin{aligned} A &= \frac{1}{n}\sum_{j=1}^{n} \left(\widehat{m}_h(t_j) - m_{\theta_0}(t_j)\right)^2 = \\ &= \frac{1}{n}\sum_{j=1}^{n} \left(\sum_{i=1}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right) Y_i - m_{\theta_0}(t_j)\right)^2 = \\ &= \frac{1}{n}\sum_{j=1}^{n} \left[\sum_{i=1}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right) (m_{\theta_0}(t_i) + \varepsilon_i) - m_{\theta_0}(t_j)\right]^2 = \\ &= A_1 + A_2 + A_3, \end{aligned} \tag{27}$$

where

$$\begin{aligned} A_1 &= \frac{1}{n}\sum_{j=1}^{n} \left[\sum_{i=1}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right) \varepsilon_i\right]^2, \\ A_2 &= \frac{1}{n}\sum_{j=1}^{n} \left[\sum_{i=1}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right) m_{\theta_0}(t_i) - m_{\theta_0}(t_j)\right]^2, \\ A_3 &= \frac{2}{n}\sum_{j=1}^{n} \left(\sum_{i=1}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right) \varepsilon_i\right)\left(\sum_{l=1}^{n} W_0^n\left(\frac{t_j - t_l}{h}\right) m_{\theta_0}(t_l) - m_{\theta_0}(t_j)\right). \end{aligned}$$

The term A_1 can be expanded into two

$$A_1 = A_{11} + A_{12} \tag{28}$$

where

$$\begin{aligned} A_{11} &= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right)^2 \varepsilon_i^2, \\ A_{12} &= \frac{1}{n}\sum_{j=1}^{n}\sum_{i\neq l}^{n} W_0^n\left(\frac{t_j - t_i}{h}\right) W_0^n\left(\frac{t_j - t_l}{h}\right) \varepsilon_i \varepsilon_l. \end{aligned}$$

It is straight forward to prove that $E(A_{12}) = 0$. Using U-statistics calculations and standard Riemann approximations it is easy to show that

$$\begin{aligned} Var(A_{12}) &= 2\sigma^4 \sum_{i \neq l} \left[\frac{1}{n} \sum_{j=1}^{n} W_0^n \left(\frac{t_j - t_i}{h} \right) W_0^n \left(\frac{t_j - t_l}{h} \right) \right]^2 \\ &\simeq 2\sigma^4 n^2 \int_0^1 \int_0^1 \varphi_n(v,w)^2 \, dv dw, \end{aligned} \tag{29}$$

where

$$\varphi_n(v,w) = \int_0^1 W_0^n \left(\frac{u-v}{h} \right) W_0^n \left(\frac{u-w}{h} \right) du. \tag{30}$$

In order to use (4) to find an asymptotic expression for (30) we need to find an asymptotic expression for the elements in the matrix $\mathbf{S}_n$. Using (5),

$$\begin{aligned} S_{n,j} &\simeq n \int_0^1 K_h(t - t_i)(t_i - t)^j \, du = n \int_{-\frac{t}{h}}^{\frac{1-t}{h}} K(v) h^j v^j dv \\ &= nh^j \int_{-\frac{t}{h}}^{\frac{1-t}{h}} v^j K(v) dv \simeq nh^j \int_{-1}^{1} v^j K(v) dv = nh^j \mu_j(K). \end{aligned}$$

For instance, an asymptotic expression for $\mathbf{S}_n$ for $p = 3$ is

$$\mathbf{S}_n \simeq \begin{pmatrix} n & 0 & nh^2\mu_2 & 0 \\ 0 & nh^2\mu_2 & 0 & nh^4\mu_4 \\ nh^2\mu_2 & 0 & nh^4\mu_4 & 0 \\ 0 & nh^4\mu_4 & 0 & nh^6\mu_6 \end{pmatrix} = n\mathbf{S},$$

where $\mathbf{S} = (s_{jl})_{1 \leq j,l \leq k}$ with $s_{jl} = h^{j+l}\mu_{j+l}(K)$ and $\mu_r(K) = \int_{-1}^{1} u^r K(u) \, du$. In the local linear case ($p = 1$) parallel calculations show that $\mathbf{S}_n \simeq n\mathbf{S}$, with

$$\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & h^2\mu_2 \end{pmatrix},$$

which, in view of (4), implies

$$W_0^n \left(\frac{u-v}{h} \right) \simeq K \left(\frac{u-v}{h} \right) \frac{1}{nh},$$

which can be easily proved for a general p. Plugging this equation in (30) and using a change of variable results in

$$\varphi_n(v,w) \simeq \frac{1}{n^2 h} \int_{-\frac{v}{h}}^{\frac{1-v}{h}} K(z) K \left(z + \frac{v-w}{h} \right) dz.$$

Now the right hand side of (29) can be written as

$$\int_0^1 \int_0^1 \varphi_n(v,w)^2 dvdw \simeq \frac{1}{n^4h^2} \int_0^1 \int_0^1 \left(\int_{-\frac{v}{h}}^{\frac{1-v}{h}} K(z) K\left(z + \frac{v-w}{h}\right) dz \right)^2 dvdw,$$

which, after standard algebra, leads to

$$\int_0^1 \int_0^1 \varphi_n(v,w)^2 dvdw \simeq \frac{1}{n^4h} \int_{-1}^1 (K * K(s))^2 ds.$$

Consequently, an asymptotic formula for the variance of A_{12} is

$$Var(A_{12}) \simeq 2\sigma^4 n^2 \frac{1}{n^4h} \int_{-1}^1 (K * K(s))^2 ds = \frac{2\sigma^4}{n^2h} \int_{-1}^1 (K * K(s))^2 ds \tag{31}$$

The expectation and variance of the term A_{11} can be easily computed

$$\begin{aligned} E(A_{11}) &= \sigma^2 \frac{1}{n} \sum_{j,i=1}^n W_0^n \left(\frac{t_j - t_i}{h}\right)^2 \\ &= \sigma^2 n \int_0^1 \int_0^1 W_0^n \left(\frac{u-v}{h}\right)^2 dudv + O\left(\frac{1}{n}\right) \\ &= \frac{\sigma^2}{nh} \int_0^1 \int_{-\frac{v}{h}}^{\frac{1-v}{h}} K(w)^2 dwdv + O\left(\frac{1}{n}\right) \\ &= \frac{\sigma^2}{nh} \int_{-1}^1 K(w)^2 dw + O\left(\frac{1}{n}\right), \end{aligned}$$

$$\begin{aligned} Var(A_{11}) &= \frac{1}{n^2} Var\left(\varepsilon_1^2\right) \sum_{i=1}^n \left[\sum_{j=1}^n W_0^n \left(\frac{t_j - t_i}{h}\right)^2 \right]^2 \\ &\simeq Var\left(\varepsilon_1^2\right) n \int_0^1 \left(\int_0^1 W_0^n \left(\frac{t-s}{h}\right)^2 dt \right)^2 ds \\ &\simeq Var\left(\varepsilon_1^2\right) \frac{n}{n^4h^4} \int_0^1 \left(\int_0^1 K \left(\frac{t-s}{h}\right)^2 dt \right)^2 ds \\ &= Var\left(\varepsilon_1^2\right) \frac{1}{n^3h^4} \left(\int K(u)^2 du \right)^2. \end{aligned}$$

Thus,

$$A_{11} = \frac{\sigma^2}{nh} \int_{-1}^1 K(w)^2 dw + O\left(\frac{1}{n}\right) + O_p\left(\frac{1}{n^{3/2}h}\right). \tag{32}$$

The term A_2 in (27) is nonrandom. It can be easily manipulated via Riemman approximations, changes of variable and Taylor expansions to conclude

$$A_2 \simeq \frac{h^4}{4} \left(\int w^2 K(w) dw \right)^2 \int_0^1 m''_{\theta_0}(u)^2 du. \tag{33}$$

On the other hand, using (31) and (32) in (28) and Condition A1, it is straight forward to prove that

$$A_1 = O_p\left(\frac{1}{n\sqrt{h}}\right).$$

which, using Cauchy-Schwarz inequality, $A_3 \leq (A_1)^{1/2}(A_2)^{1/2}$, and expression (33) yields

$$A_3 = O_p\left(n^{-\frac{1}{2}}h^{\frac{5}{4}}\right). \tag{34}$$

As a consequence, using Condition A1 and expressions (33 and (34) it is easy to prove that the last two terms in (27) are negligible for the purpose of the asymptotic distribution in (33):

$$\sqrt{n^2h}A_2 = o_P(1)\,,\ \sqrt{n^2h}A_3 = o_P(1)\,. \tag{35}$$

Now, in view of Condition A1, (27), (35), (28) and (32), a sufficient condition for (26) is

$$\sqrt{n^2h}A_{12} \xrightarrow{d} N\left(0, 2\sigma^4\int_{-1}^{1}(K*K)^2\right). \tag{36}$$

The proof of (36) is omitted here since it follows the lines in GONZÁLEZ-MANTEIGA and CAO (1993), using DE JONG (1987) Central Limit Theorem for generalized quadratic forms.

References

[1] ALCALÁ, J.T., CRISTÓBAL, J.A. and GONZÁLEZ-MANTEIGA, W. (1999). Goodness-of-fit test for linear models based on local polynomials. Statistics & Probability Letters, **42**, 39-46.

[2] CAO, R., NAYA, S., ARTIAGA, R., GARCIA, A. and VARELA, A. (2004). Logistic approach to polymer degradation in dynamic TGA. *Polymer Degradation and Stability*, **85**, 667-674.

[3] DETTE H. and MUNK, A. (1998). Validation of linear regression models. *Annals of Statistics*, **26**, 778-800.

[4] DIEBOLT, J. and ZUBER, J. (2001). On testing the goodness-of-fit of nonlinear heteroscedastic regression models. *Communications in Statistics – Simulations and Computation*, **30**, 195-216.

[5] ESCANCIANO, J.C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, **22**, 1030-1051.

[6] FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications.* Chapman and Hall. London.

[7] GONZÁLEZ-MANTEIGA, W. and CAO, R. (1993). Testing hypothesis of general linear model using non-parametric regression estimators. *Test*, **2**, 161-189.

[8] GONZÁLEZ-MANTEIGA, W. and PÉREZ-GONZÁLEZ, A. (2006). Goodness-of-fit tests for linear regression models with missing response data *Canadian Journal of Statistics*, **34**, 149-170.

[9] GONZÁLEZ-MANTEIGA, W. and VILAR, J.M. (1995). Testing linear regression models using non-parametric regression estimators when errors are non-independent. *Computational Statistics and Data Analysis*, **20**, 521-541.

[10] HÄRDLE, W. and KNEIP, A. (1999). Testing a regression model when we have smooth alternatives in mind. *Scandinavian Journal of Statistics*, **26**, 221-238.

[11] HÄRDLE, W. and MAMMEN, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, **21**, 1926-1947.

[12] HÄRDLE, W., MAMMEN, E. and MÜLLER, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association*, **93**, 1461-1474.

[13] JONG, P. DE (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, **75**, 261-277.

[14] KAUERMANN, G. and TUTZ G. (2001). Testing generalized linear and semiparametric models against smooth alternatives. *Journal of the Royal Statistical Society Series B*, **63**, 147-166.

[15] KHMALADZE, E.V. and KOUL, H.L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Annals of Statistics*, **32**, 995-1034.

[16] LI Q. and WANG S.J. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, **87**, 145-165.

[17] LIN, D.Y., WEI, L.J. and YING, Z. (2002). Model-checking techniques based on cumulative residuals, *Biometrics*, **58**, 1-12.

[18] NAYA, S. (2003). *Nuevas aplicaciones de la estimación paramétrica y no paramétrica de curvas al Análisis Térmico*. Doctoral Thesis. Universidade da Coruña.

[19] NAYA, S., CAO, R. and ARTIAGA, R. (2003). Local polynomial estimation of TGA derivatives using logistic regression for pilot bandwidth selection. *Thermochimica Acta.* **6**, 319-322.

[20] NAYA, S., CAO, R., LÓPEZ-DE-ULLIBARRI, I., ARTIAGA, R., BARBADILLO, F. and GARCIA, A. (2007). Logistic mixture model vs Arrhenius for kinetic study of degradation of materials by dynamic thermogravimetric analysis. To appear in *Journal of Chemometrics*.

[21] NEUMEYER, N. and DETTE, H. (2003). Nonparametric comparison of regression curves: An empirical process approach. *Annals of Statistics*, **31**, 880-920.

[22] PETROV V. (1995). *Limit theorems of probability theory. Oxford Studies in Probability*. Oxford university Press.

[23] RODRÍGUEZ-CAMPOS, M.C., GONZÁLEZ-MANTEIGA, W. and CAO, R. (1998). Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation. *Journal of Statistical Planning and Inference*, **67**, 99-122.

[24] RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association,* **90**, 1257-1267.

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

[26] STUTE, W. and GONZÁLEZ-MANTEIGA, W. (1996). NN goodness-of-fit test for linear models. *Journal of Statistical Planning and Inference*, **53**, 75-92.

[27] STUTE, W, GONZÁLEZ-MANTEIGA, W. and PRESEDO-QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, **93**, 141-149.

[28] STUTE, W., THIES, S. and ZHU, L.X. (1998). Model checks for regression: An innovation process approach. *Annals of Statistics*, **26**, 1916-1934.

[29] STUTE, W. and ZHU, L.X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics*, **29**, 535-545.

[30] STUTE, W. and ZHU, L.X. (2005). Nonparametric checks for single-index models. *Annals of Statistics*, **33**, 1048-1083.

[31] VILAR, J.M. and GONZÁLEZ-MANTEIGA, W. (2000). Resampling for checking linear regression models via non-parametric regression estimation. *Computational Statistics and Data Analysis*, 35, 211-231.

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

[33] ZHU, L.X. (2003). Model checking of dimension-reduction type for regression. *Statistica Sinica*, **13**, 283-296.