Testing generalized linear and semiparametric models against smooth alternatives

Göran Kauermann, Gerhard Tutz

Institutions: University of Glasgow, Ludwig Maximilian University of Munich

Published on: 01 Jan 2001 - Journal of The Royal Statistical Society Series B-statistical Methodology (Blackwell Publishers Ltd.)

Topics: Semiparametric regression, Semiparametric model, Test statistic, Generalized additive model and Asymptotic distribution

Kauermann, Tutz:

Testing Generalized Linear and Semiparametric Models Against Smooth Alternatives

Sonderforschungsbereich 386, Paper 149 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

Testing Generalized Linear and Semiparametric Models Against Smooth Alternatives

Göran Kauermann and Gerhard Tutz

Ludwig-Maximilians-University Munich
Akademiestraße 1
80799 München

kauerman@stat.uni-muenchen.de tutz@stat.uni-muenchen.de

Abstract

We propose goodness of fit tests for testing generalized linear models and semiparametric regression models against smooth alternatives. The focus is on models having both, continuous and factorial covariates. As smooth extension of a parametric or semiparametric model we use generalized varying coefficient models as proposed by Hastie & Tibshirani (1993). A likelihood ratio statistic is used for testing, and asymptotic normality of the test statistic is proven. Due to a slow asymptotic convergence rate a bootstrap approach is pursued. Asymptotic expansions allow to write the estimates as linear smoothers which in turn guarantees simple and fast bootstrapping. The test is shown to have $\sqrt{n}$ power, but in contrast to parametric tests it is powerful against smooth alternatives in general.

Keywords: Likelihood Ratio, Local Likelihood Fitting, Model Checking, Semiparametric Models, Smoothing

1 Introduction

In recent years several articles dealt with goodness of fit tests for checking parametric models against smooth alternatives. The focus has been on testing a generalized linear regression model of the form

$$H_0 \quad : \quad E(y|u) = h\{V(u)\beta\} \tag{1}$$

with continuous regressors u, design matrix $V(u)$ and known link function $h(\cdot)$ against the smooth mode

$$H_1 \quad : \quad E(y|u) = h\{\gamma(u)\}, \tag{2}$$

where $\gamma(u)$ is an unknown but smooth function in u. Assuming that $V(u)$ consists of smooth but known functions in u, model (1) is a proper submodel of (2). A typical example is the case where $V(u)$ consists of polynomials in u. Smooth tests for testing (1) against (2) can be derived by smoothing the fitted parametric residuals from (1), see e.g. le Cessie & van Houwelingen (1991) or Azzalini & Bowman (1993). Alternatively one can compare the parametric and the smooth models using a likelihood ratio type statistic, see for instance Azzalini, Bowman & Härdle (1989), Staniswalis & Severini (1991) or Härdle & Mammen (1993). Another approach was suggested by Firth, Glosup & Hinkley (1991) who estimate the parametric model locally and take the improvement of the fit as a measure for goodness of fit. Eubank & Spiegelman (1990), Eubank, Hart & LaRiccia (1993), Hart & Wehrly (1992) or Aerts, Claeskens & Hart (1998) extend $V(u)$ by appropriately chosen basis functions and assess whether the additional model components improve the fit significantly. Recently, Stute (1997) suggested a test based on integrated regression functions, while Dette & Munk (1998) extend nonparametric tests for testing heteroscedasticity in regression models. Further approaches for model checking have been suggested among others by Raz (1990), Müller (1992) or Kauermann & Tutz (1998b). A comprehensive overview of smooth tests can be found in Hart (1997).

In contrast to most of the papers cited above we consider models with both, factorial covariates x, say, and continuous regressors u. A parametric model that jointly includes continuous and factorial regressors is the generalized linear model

$$E(y|x,u) \quad = \quad h\{W(x,u)\beta\} \tag{3}$$

where the design matrix $W(x,u)$ is constructed from both, x and u. The smooth alternative corresponding to (3) is a varying coefficient model in the sense of Hastie & Tibshirani (1993), which is given by

$$E(y|x,u) \quad = \quad h\{Z(x)\gamma(u)\}. \tag{4}$$

Matrix $Z(x)$ is a design matrix built solely from the factorial regressors x and $\gamma(u)$ is a vector valued smooth but unknown function. For instance, if x is a binary factor, a smooth alternative to the parametric linear interaction model $E(y|x,u) = h(\beta_0 + u\,\beta_u + x\,\beta_x + u\,x\,\beta_{ux})$ is the model $E(y|x,u) = h\{\gamma_0(u) + x\gamma_x(u)\}$. Here $\gamma_0(u)$ is the smooth main effect and $\gamma_x(u)$ is the effect of x modified by u, i.e. the smooth interaction between x and u. We consider multivariate parametric interaction models by assuming that the design matrix $W(x,u)$ in (3) decomposes into the matrix product $W(x,u) = Z(x)V(u)$. This ensures that the parametric model is a proper submodel of the varying coefficient model (4), since $\gamma(u)$ is modeled parametrically by $V(u)\beta$. For identifiability reasons we further assume that $V(u)$ has full rank and has a row diagonal structure. This means in each column of $V(u)$

there is only a single non-zero element. Moreover, to ensure that models are nested each row of $V(u)$ is assumed to have 1 as element. For instance the parametric model $E(y|x,u) = h(\beta_0 + u\,\beta_u + x\,\beta_x + u\,x\,\beta_{ux})$ may be written as

$$E(y|x,u) = h\left\{(1,x)\begin{pmatrix} 1 & u & 0 & 0 \\ 0 & 0 & 1 & u \end{pmatrix}\beta\right\}$$

with $\beta^T = (\beta_0, \beta_u, \beta_x, \beta_{xu})$ and obvious definition for $Z(x)$ and $V(u)$. If the polynomial degrees in the rows of $V(u)$ coincide, which is the case in this example, we can also write $V(u)$ as Kronecker product $V(u) = I \otimes (1,u)$ with I as identity matrix. In general however the polynomial degree in the rows of $V(u)$ is allowed to differ.

We propose tests for testing the H_0 model (3) against the alternative (4). In order to avoid the disturbing influence of the smoothing bias (see e.g. Härdle & Mammen 1993) and to allow for appropriate bandwidth selection we estimate the alternative model (4) by locally fitting the parametric model (3). This implies that under H_0 the smooth estimates are estimated without the typical smoothing bias, so that bias consideration can be neglected. Moreover, in a first order expansion the estimate is obtained by linear smoothing. In particular this provides simple and numerically fast calculation of the fit .

A related but different testing problem occurs in partial linear or semiparametric models where effects of the continuous and factorial regressors are modeled additively by

$$E(y|x,u) = h\{\gamma_0(u) + Z_x(x)\beta_x\}, \tag{5}$$

see for instance Heckman (1986), Speckman (1988), Severini & Staniswalis (1994) or Hunsberger (1995). In (5), the design matrix $Z_x(x)$ is built from x however without the intercept in order to ensure identifiability in (5). The covariates x and u do not interact, i.e. the smooth effect of the continuous variable $\gamma_0(u)$ is shifted for different values of the factorial variables x. Härdle, Mammen & Müller (1998) propose tests for testing whether the shape of $\gamma_0(u)$ can be modeled parametrically (see also Fan & Li 1996). Bowman & Young (1996) investigate nonparametrically whether smooth main effects differ in factorial groups, i.e whether $\beta_x \equiv 0$. Our focus is on testing the model assumption that x and u act additively.

We fit the parameters in the semiparametric model (5) by a combination of profile likelihood and local likelihood estimation (see Cuzick 1992 or Severini & Wong 1992). For normal response and identity link this approach is equivalent to Speckman's (1988) estimate. For testing purposes, the varying coefficient model (4) is considered as smooth alternative to (5), with $Z(x) = \{1, Z_x(x)\}$ in (4). The alternative model is again estimated by locally fitting the H_0 model (5). This means locally a semiparametric model is fitted which in turn allows the effects of the factors to vary. The welcome benefit of this estimation approach is that bias components of the smooth fit cancel out and hence the typical smoothing bias can again be neglected.

For both settings, i.e. for parametric and semiparametric models we employ a likelihood ratio statistic. Asymptotic normality is proven with convergence rate of order $O(\lambda^{1/2})$, where λ is the bandwidth of the smooth fit with $\lambda \to 0$. The asymptotic rate of convergence is rather slow so that a bootstrap approach is pursued. Asymptotic approximations are used to provide simple and numerically fast computation. The proposed test is shown to be asymptotically as powerful as classical parametric likelihood ratio tests. This means it detects general but smooth alternatives tending to H_0 with order $\sqrt{n}$. In contrast to parametric tests however the smooth test has an omnibus power which also shows in simulations.

2 Testing Generalized Linear Models

2.1 Local Likelihood Fitting

Let the response y for given x and u follow the exponential family distribution $y|x,u \sim \exp[\{y\theta - \kappa(\theta)\}/\phi]$, where $\theta = \theta(\mu)$ is the natural parameter, $\mu = E(y|x,u)$ is the expectation and $\kappa(\theta)$ is the log normalization constant. The dispersion parameter ϕ is either assumed to be known or taken as nuisance parameter. Let (y_i, x_i, u_i) denote a random sample for $i = 1, \dots, n$ and abbreviate $Z_i = Z(x_i)$, $V_i = V(u_i)$ and $W_i = Z_i V_i$. In the following the objective is to test the generalized linear model

$$H_0 : \quad E(y|x,u) \quad = \quad h\{Z(x)V(u)\beta\} \tag{6}$$

against the varying coefficient model

$$H_1 : \quad E(y|x,u) \quad = \quad h\{Z(x)\gamma(u)\}. \tag{7}$$

The varying coefficient $\gamma(u)$ under H_1 is estimated by local likelihood (see e.g. Fan, Heckman & Wand 1995 or Carroll, Ruppert & Welsh 1988). Having in mind that H_0 is to be investigated, the local likelihood is based on the H_0 model. This means we fit the H_0 model (6) locally by introducing kernel weights $\omega_{\lambda,ij} = K\{(u_i - u_j)/\lambda\}/K(0)$ with $K(\cdot)$ as unimodal kernel function and λ as smoothing parameter. For $u = u_i$ this yields the local likelihood function

$$\mathbf{l}_{(i)}(\beta_i) \quad = \quad \sum_j \omega_{\lambda,ij} l_j(W_j \beta_i) \tag{8}$$

where $l_j(\eta) = y_j\theta - \kappa(\theta)$ with $\theta = \theta\{h(\eta)\}$ is the log likelihood contribution of the jth observation evaluated at the linear predictor η. Maximizing (8) with respect to β_i yields the local likelihood estimate $\widehat{\gamma}_i = \widehat{\gamma}(u_i) = V_i\widehat{\beta}_i$. If the matrix $V(u)$ consists of polynomials in u, estimates of this type are also known as local polynomial estimates, see e.g. Fan & Gijbels (1996).

When investigating the asymptotic properties of estimates obtained from (8) we assume standard regularity conditions. For instance we postulate that $\gamma(u)$ is sufficiently smooth and that locally weighted Fisher matrices have full rank, see Kauermann & Tutz (1998a) for a technical discussion of these assumptions. Differentiating (8) with respect to β leads to the local estimating equation

$$0 \quad = \quad \sum_j \omega_{\lambda,ij} W_j^T l_{\eta,j}(W_j\widehat{\beta}_i), \tag{9}$$

where $l_{\eta,j}(\eta) = \partial l_j(\eta)/\partial\eta = \{\partial h(\eta)/\partial\eta\}\mathrm{var}(y_j)^{-1}\{y_j - h(\eta)\}$ is the standard score contribution. As shown in the appendix, expansion of (9) yields in first order approximation

$$\widehat{\gamma}_i - \gamma_i \quad \approx \quad V_i\mathbf{F}_{(i)}^{-1}\left\{\sum_j \omega_{\lambda,ij} W_j^T l_{\eta,j}(\eta_j)\right\} + V_i b_{(i),\lambda} \tag{10}$$

where $\eta_j = Z_j\gamma_j$ is the true predictor and $\mathbf{F}_{(i)} = \sum_j \omega_{\lambda,ij} W_j^T F_j W_j$ is the locally weighted Fisher matrix with $F_j = F(\eta_j) = E\{-\partial^2 l(\eta_j)/(\partial\eta)^2\}$. The component $b_{(i),\lambda}$ contains the smoothing bias which equals

$$b_{(i),\lambda} \quad = \quad \mathbf{F}_{(i)}^{-1}\left\{\sum_j \omega_{\lambda,ij} W_j^T F_j Z_j(\gamma_j - \gamma_i)\right\}.$$

It is useful to give expansion (10) in matrix notation. Let S_λ denote the $n \times n$ dimensional *generalized smoothing matrix* with entries

$$S_{\lambda;ij} \quad = \quad \omega_{\lambda,ij} W_i \mathbf{F}_{(i)}^{-1} W_j^T. \tag{11}$$

Note that the rows of $S_\lambda\,\text{Diag}(F_i)$ sum up to one, with $\text{Diag}(F_i)$ denoting the diagonal matrix with F_i, $i = 1, \ldots, n$, as diagonal elements. Let in the sequel $\eta = (\eta_1, \ldots, \eta_n)^T = (Z_1\gamma_1, \ldots, Z_n\gamma_n)^T$ be the vector of predictors and $\widehat{\eta} = (Z_1\widehat{\gamma}_1, \ldots, Z_n\widehat{\gamma}_n)^T$ be the corresponding estimate. Moreover, let $l_\eta = (l_{\eta,1}, \ldots, l_{\eta,n})^T$ be the score vector with $l_{\eta,i} = l_{\eta,i}(\eta_i)$ as elements. From (10) one obtains

$$\widehat{\eta} - \eta \;\approx\; S_\lambda l_\eta + B_\lambda \tag{12}$$

where the bias $B_\lambda = (W_1 b_{(1),\lambda}, \ldots, W_n b_{(n),\lambda})$ equals $B_\lambda = S_\lambda \text{Diag}(F_i)\eta - \eta$. If model H_0 holds we have $\eta_j = W_j\beta$ which provides $S_\lambda\text{Diag}(F_i)\eta = \eta$. Hence under H_0 the smoothing bias B_λ vanishes.

Local likelihood fitting based on solving (9) typically demands time consuming computation since locally iterative fitting is required. This can be avoided by making use of the fit under the H_0 model. Let $P = W(W^T\text{Diag}(F_i)W)^{-1}W^T$ with $W^T = (W_1^T, \ldots, W_n^T)$ be the projection type matrix resulting from fitting the H_0 model by standard maximum likelihood. This means under H_0 one has the first order approximation $\widehat{\eta}^{(0)} - \eta \approx Pl_\eta$ where $\widehat{\eta}^{(0)} = W\widehat{\beta}$ with $\widehat{\beta}$ as maximum likelihood estimate under the H_0 model. When fitting η under H_1 one can employ expansion (12) but substitute the unknown predictor η by the fit under H_0. This means we define the one step estimate $\widehat{\eta}^{(1)} := \widehat{\eta}^{(0)} + S_\lambda\widehat{l}_\eta^{(0)}$ with $\widehat{l}_\eta^{(0)} = \{l_{\eta,1}(\widehat{\eta}_1^{(0)}), \ldots, l_{\eta,n}(\widehat{\eta}_n^{(0)}\}$ denoting the fitted score vector. Making use of $S_\lambda\text{Diag}(F_i)\widehat{\eta}^{(0)} = \widehat{\eta}^{(0)}$ and expanding $\widehat{l}_\eta^{(0)}$ about η gives in first order approximation

$$\widehat{\eta}^{(1)} - \eta \;\approx\; S_\lambda l_\eta + B_\lambda.$$

This means that $\widehat{\eta}^{(1)}$ equals in first order approximation the local likelihood estimate $\widehat{\eta}$ defined in (9). In contrast to the local likelihood estimate however, $\widehat{\eta}^{(1)}$ is calculated as linear smoother in one step, starting from the fit under H_0, and hence provides simple and fast calculation.

2.2 Likelihood Ratio Testing

We test the H_0 model against the alternative H_1 by use of the likelihood ratio statistic

$$\Lambda_\lambda = -2 \sum_i \{l_i(\widehat{\eta}_i^{(0)}) - l_i(\widehat{\eta}_i^{(1)})\} \tag{13}$$

where subscript λ indicates the dependence on the smoothing parameter. In first order approximation under H_0 the likelihood ratio is approximated by

$$\begin{aligned} \Lambda_\lambda &= 2\sum_i l_i(\widehat{\eta}_i^{(1)}) - l_i(\widehat{\eta}_i^{(0)}) \\ &\approx 2l_\eta^T(\widehat{\eta}^{(1)} - \eta) - (\widehat{\eta}^{(1)} - \eta)^T \mathrm{Diag}(F_i)(\widehat{\eta}^{(1)} - \eta) \\ &\qquad -2l_\eta^T(\widehat{\eta}^{(0)} - \eta) + (\widehat{\eta}^{(0)} - \eta)^T \mathrm{Diag}(F_i)(\widehat{\eta}^{(0)} - \eta) \\ &\approx l_\eta^T\{2S_\lambda - {S_\lambda}^T \mathrm{Diag}(F_i) S_\lambda - P\} l_\eta. \end{aligned} \tag{14}$$

where we made use of the property $P = P\mathrm{Diag}(F_i)P$. If model H_0 holds, efficient estimation of $\gamma(u)$ is achieved only for the unsmoothed case $\lambda \to \infty$. Under H_1, however, the usual rate for (univariate) smoothing is $\lambda \to 0$ and $\lambda n \to \infty$, which is assumed in the following. As shown in the appendix, the quadratic form (14) allows to easily calculate the moments of Λ_λ. With $\tilde{S}_\lambda = \{2S_\lambda - S_\lambda^T \mathrm{Diag}(F_i) S_\lambda\}\mathrm{Diag}(F_i)$ one obtains in first order approximation $E_{H_0}(\Lambda_\lambda) \approx \mathrm{tr}(\tilde{S}_\lambda) - q$, where q is the rank of W. The term $\mathrm{tr}(\tilde{S}_\lambda)$ thereby is frequently called the degree of freedom for smoothing (see Hastie & Tibshirani 1990). Cornish-Fisher expansion (see e.g. Barndorff-Nielsen

& Cox 1989) leads to

$$P\left(\frac{\Lambda_\lambda - E(\Lambda_\lambda)}{\sqrt{\mathrm{Var}(\Lambda_\lambda)}} \le z\right) \quad = \quad \Phi(z) - \phi(z)\frac{\mathrm{Cum}_3(\Lambda_\lambda)}{6\mathrm{Var}(\Lambda_\lambda)^{3/2}}(z^2-1) + \dots \qquad (15)$$

where $\Phi()$ and $\phi()$ denote the distribution and density function of a standard normal distribution. For $\lambda \to 0$ the cumulants of Λ_λ tend to infinity with order λ^{-1}. This implies that $\mathrm{Cum}_3(\Lambda_\lambda) = O(\lambda^{-1})$ and $\mathrm{Var}^{-1}(\Lambda_\lambda) = O(\lambda)$, as demonstrated in the appendix. The latter component in (15) tends to zero with order $O(\lambda^{1/2})$ and components not explicitly listed are $O(\lambda)$. Hence for $\lambda \to 0$ the likelihood ratio Λ_λ is asymptotically normal, however the rate of convergence is rather slow. Therefore a bootstrap procedure seems more appropriate for testing purposes. We suggest to bootstrap directly from (14), i.e.

$$\Lambda^*_\lambda = l^{*T}_\eta \{2S_\lambda - S^T_\lambda \mathrm{Diag}(F_i) S_\lambda - P_\beta\} l^*_\eta \qquad (16)$$

where $l^{*T}_\eta = (l^*_{\eta,1}, \dots, l^*_{\eta,n})$ with $l^*_{\eta,i} = \{\partial h(\hat\eta^{(0)}_i)/\partial\eta\}\mathrm{var}(y_i)^{-1}\{y^*_i - h(\hat\eta^{(0)}_i)\}$ and y^*_i drawn from the fitted parametric model with predictor $\hat\eta^{(0)}_i = W_i\hat\beta$.

Power Consideration

We briefly discuss the power properties of the test. We consider alternatives of the type $H_1 : \gamma(u) = V(u)\beta + \varphi(u)n^{-p}$, with $p > 0$ and $\varphi(u)$ being some arbitrary but smooth function, bounded and bounded away from zero, i.e. $0 < a \le \sum_j^n |\varphi(u_j)|/n \le b < \infty$. Moreover $\varphi(u)$ is assumed to be identifiable, i.e. $\varphi(u)$ and $V(u)$ are orthogonal as explicitly stated in the appendix. It is shown in the appendix that for $p < 1/2$, the test detects H_1 asymptotically with probability one. Hence one achieves the same order of power as typically met in standard parametric settings. However in contrast to parametric tests, $\varphi()$ is arbitrary but smooth here and therefore the test has sensible power for general smooth alternatives.

Choice of the bandwidth

We suggest choosing λ from the Akaike criterion

$$\widehat{\lambda} := \arg\max\{\Lambda_\lambda - 2\, E_{H_0}(\Lambda_\lambda)\}, \tag{17}$$

where it is advisable to restrict the range of λ in order to avoid undersmoothing. For instance one can restrict the degree of freedom of the H_1 model to exceed the parametric degree only by a certain amount. In the simulation and example below we set $\text{tr}(\tilde{S}_\lambda) \leq q + 1$ with q as parametric degree of freedom. In general it can be observed that the significance of Λ_λ depends only weakly on the bandwidth λ, i.e. the p value changes rather moderately for different bandwidths. The major reason for this property is that due to fitting the H_0 model locally the smoothing bias disappears under H_0 and in the extreme case of smoothing, i.e. $\lambda \to \infty$, the fits of H_1 and H_0 coincide.

2.3 Simulation and Example

Simulation Study:

In a simulation study the main effect logit model $H_0 \;:\; E(y|x,u) = \text{logit}^{-1}(\beta_0 + u\beta_u + x\beta_x)$ with a balanced binary factor x is tested. The covariate u takes 30 equidistant points in $[0,1]$ and at each point of u five repetitions of the binary response y are sampled at $x = -1$ and $x = 1$ with the predictor given by $\eta = -0.5 + u + x$. The power of the test is assessed by simulating from the alternative models $H_{1,a} : \eta = -1 + u + x - xu$; $H_{1,b} : \eta = -1 + u - x\{0.5 - \sin(u\pi)\}$ and $H_{1,c} : \eta = -2u + 2u^2 + 0.5 * xu$. Table 4 shows the simulated rejection frequencies based on 500 simulations, each one based on 1000 bootstraps replicates. For comparison we also report the rejection probabilities of a parametric likelihood ratio test obtained from testing the H_0 model against the parametric interaction model $H_1 : \eta = \beta_0 + u\beta_u + x\beta_x + x\,u\,\beta_{xu}$. The smooth test behaves slightly liberal but shows omnibus power by indicating lack of fit in all three alternative settings. In contrast, the parametric test shows power only for model $H_{1,a}$, which is the correct alternative model in the likelihood ratio. In settings $H_{1,b}$ and $H_{1,c}$ however the power of the parametric test is disappointing.

(Table 1)

Example:

We investigate a dataset given in Bowman & Azzalini (1997). The data describe the occurrence of human parasitic worm infestation ($y = 1$ for yes, 0 for no) of $n = 304$ citizens of a rural community in China. The explanatory quantities are age, u, and gender, x. We test the main effect logit model $E(y|x, u) = \text{logit}^{-1}(\beta_0 + x\beta_x + u\beta_u)$ against the smooth alternative $\text{logit}^{-1}\{\gamma_0(u) + x\gamma_x(u)\}$ yielding a p-value of 0.03 with $\widehat{\lambda} = 30$ chosen by (17). Figure 1 shows the parametric fit and the corresponding fit under H_1. We extend the parametric model by a linear interaction term for u and x which provides the p-value 0.09 at $\widehat{\lambda} = 55$. Modeling an additional quadratic main effect for age finally gives the p-value 0.35 at $\widehat{\lambda} = 70$. Hence, the quadratic interaction model $E(y|x, u) = \text{logit}^{-1}(\beta_0 + x\beta_x + xu\beta_{xu} + u\beta_u + u^2\beta_{uu})$ can be considered as an adequate model for the data. This is also seen from Figure 1 where nonparametric and parametric fits hardly differ.

(Figure 1)

3 Testing Semiparametric Models

3.1 Profile Likelihood Fitting

In the following section we extend the above testing problem by considering the semiparametric model

$$H_0 \; : \quad E(y|x,u) \;\; = \;\; h\{\gamma_0(u) + Z_x(x)\beta_x\}, \tag{18}$$

which is tested against the varying coefficient model

$$H_1 \; : \quad E(y|x,u) = h\{Z(x)\gamma(u)\} \tag{19}$$

with $Z(x) = \{1, Z_x(x)\}$. In the semiparametric model the regressors x and u act additively, i.e. $\gamma_0(u)$ is the smooth main effect and $Z_x(x)\beta_x$ is an additive shift for the factors. Hence testing (18) against (19) is a test on interaction between the factorial covariates x and the continuous regressors u.

Estimation of the semiparametric model (18) requires both, local likelihood fitting for the smooth component and profile likelihood fitting for the parametric component β_x, see Severini & Wong (1992) or Hunsberger (1995). We first consider β_x as known so that $\eta_x = Z_x(x)\beta_x$ serves as given offset in the smooth model $E(y|x,u) = h\{\gamma_0(u) + \eta_x\}$. We estimate $\gamma_0(u)$ by fitting locally the model $E(y|x,z) = h\{V_0(u)\beta_0 + \eta_x\}$ where $V_0(u)$ is a row vector of polynomials in u. For instance the choice $V_0(u) = (1, u)$ corresponds to local linear fitting of $\gamma_0(u)$. Solving the estimating equation

$$0 \;\; = \;\; \sum_j \omega_{\lambda,ij} V_{0,j}^T l_{\eta,j}(V_{0,j}\widehat{\beta}_{0,i} + \eta_{x,j}) \tag{20}$$

yields the estimate $\widehat{\gamma}_{0,i} = V_{0,i}\widehat{\beta}_{0,i}$ for the smooth component, where $V_{0,j} = V_0(u_j)$. One should note that $\widehat{\gamma}_{0,i}$ calculated from (20) depends on the particular value of parameter β_x, which is however suppressed in the notation. Now the estimates $\widehat{\gamma}_{0,i}$ are inserted into the likelihood for β_x yielding the profile likelihood function $\sum_i l_i(\widehat{\gamma}_{0,i} + Z_{x,i}\beta_x)$. Differentiating this profile likelihood with respect to β_x gives the estimating equation for β_x

$$0 \;=\; \sum_i \tilde{Z}_{x,i}^T l_{\eta,i}(\widehat{\gamma}_{0,i} + Z_{x,i}\widehat{\beta}_x) \tag{21}$$

with $\tilde{Z}_{x,i} = Z_{x,i} + (\partial\widehat{\gamma}_{0,i}/\partial\beta_x^T)$. The derivative $\partial\widehat{\gamma}_{0,i}/\partial\beta_x^T$ can be calculated by differentiating (20) with respect to β_x. As shown in the appendix in first order approximation one obtains

$$\begin{aligned} \frac{\partial\widehat{\gamma}_{0,i}}{\partial\beta_x^T} &\approx -V_{0,i}\left\{\sum_j \omega_{\lambda,ij}V_{0,j}^T F_j V_{0,j}\right\}^{-1}\left\{\sum_j \omega_{\lambda,ij}V_{0,j}^T F_j Z_{x,j}\right\} \\ &= -S_{0,\lambda;i\bullet}\mathrm{Diag}(F_i)Z_x \end{aligned} \tag{22}$$

with $F_j = F(\eta_j)$, $Z_x = (Z_{x,1}^T, \ldots Z_{x,n}^T)^T$ and $S_{0,\lambda;i\bullet}$ denoting the ith row of the generalized smoothing matrix $S_{0,\lambda}$ with entries

$$S_{0,\lambda;ij} \;=\; \omega_{\lambda,ij}V_{0,i}\left\{\sum_r \omega_{\lambda,ir}V_{0,r}^T F_r V_{0,r}\right\}^{-1} V_{0,j}^T. \tag{23}$$

For normally distributed response and identity link this estimation procedure was first suggested by Speckman (1988). Asymptotic investigation of the two estimating equations (20) and (21) allows to rewrite the fit in matrix notation. Let $\tilde{P} = \tilde{Z}_x(\tilde{Z}_x^T\mathrm{Diag}(F_i)\tilde{Z}_x)^{-1}\tilde{Z}_x^T$ be the projection type matrix for fitting the parametric component where $\tilde{Z}_x = Z_x - S_{0,\lambda}\mathrm{Diag}(F_i)Z_x$. We show in the appendix, that in first order approximation, the estimated predictor $\widehat{\eta}_i^{(0)} = \widehat{\gamma}_{0,i} + Z_{x,i}\widehat{\beta}_x$ obtained from the above routine fulfills

$$\widehat{\eta}^{(0)} - \eta \approx (S_{0,\lambda} + \tilde{P}\tilde{I}_\lambda)l_\eta + B_{0,\lambda} \tag{24}$$

where $\tilde{I}_\lambda = \{I - \text{Diag}(F_i)S_{0,\lambda}\}$ and l_η denotes the score vector. The component $B_{0,\lambda}$ contains the bias due to smoothing which equals $B_{0,\lambda} = S_{0,\lambda}\text{Diag}(F_i)\gamma_0 - \gamma_0$ where $\gamma_0 = (\gamma_{0,1}, \dots \gamma_{0,n})^T = \{\gamma_0(u_1), \dots, \gamma_0(u_n)\}^T$.

The next step is to consider estimation under the alternative model (19). This could in general be done by a local likelihood approach as suggested in the previous section. However, for testing purposes it seems more natural to fit the H_1 model by locally fitting the H_0 model, a concept which has also been pursued in the previous section. The main advantage there has been that due to local fitting of the H_0 model the smoothing bias could be neglected. We show now that the same property also holds for the semiparametric setting. Our intention is to construct an estimate which can be seen as a smooth version of the estimate $\widehat{\eta}^{(0)}$ found in the H_0 model. This can be achieved by substituting the parametric projection matrix $\tilde{P}$ in (24) by a corresponding smooth version. Let therefore $\tilde{S}_{x,\mu}$ denote the generalized smoothing matrix with entries

$$\tilde{S}_{x,\mu;ij} \quad = \quad \omega_{\mu,ij}\tilde{Z}_{x,i}\left(\sum_k \omega_{\mu,ik}\tilde{Z}_{x,k}^T F_k \tilde{Z}_{x,k}\right)^{-1}\tilde{Z}_{x,j}^T \tag{25}$$

where $\omega_{\mu,ij}$ denote some kernel weights with μ as a second bandwidth. One should note that for $\mu \to \infty$ matrix $\tilde{S}_{x,\mu}$ becomes $\tilde{P}$. For $\mu \to 0$ on the other hand the resulting fit uncovers the smooth structure which is not modeled in the semiparametric model. Substituting $\tilde{P}$ in (24) by $\tilde{S}_{x,\mu}$ suggests the one step estimate

$$\widehat{\eta}^{(1)} := \widehat{\eta}^{(0)} + (S_{0,\lambda} + \tilde{S}_{x,\mu}\tilde{I}_\lambda)\widehat{l}_\eta^{(0)} + \widehat{B} \tag{26}$$

where $\widehat{l}_\eta^{(0)} = \{l_{\eta,1}(\widehat{\eta}_1^{(0)}), \dots, l_{\eta,n}(\widehat{\eta}_n^{(0)})\}^T$ and $\widehat{B}$ as estimated bias defined by $\widehat{B} = (S_{0,\lambda} + \tilde{S}_{x,\mu}\tilde{I}_\lambda)\text{Diag}(F_i)\widehat{\eta}^{(0)} - \widehat{\eta}^{(0)}$. In the appendix it is shown that under H_0 in first order approximation one obtains

$$\widehat{\eta}^{(1)} \approx \eta + (S_{0,\lambda} + S_{x,\mu}\tilde{I}_\lambda)l_\eta + B_{0,\lambda}. \tag{27}$$

This shows that $\widehat{\eta}^{(1)}$ and $\widehat{\eta}^{(0)}$ have the same first order smoothing bias under H_0. Moreover, for $\mu \to \infty$ both fits coincide in first order approximation, i.e. $\widehat{\eta}^{(1)} \to \widehat{\eta}^{(0)}$. The bandwidth μ steers the additional structure in the fit of model H_1 compared to model H_0, while bandwidth λ controls the smoothness of the main effect only. Hence, μ is of primary interest for testing purposes.

3.2 Likelihood Ratio Testing

We test the semiparametric model (18) against the smooth alternative (19) using the likelihood ratio

$$\Lambda_\mu = -2\sum_i \{l_i(\widehat{\eta}_i^{(0)}) - l_i(\widehat{\eta}_i^{(1)})\} \tag{28}$$

where the subscript μ here indicates the dependence on the smoothing parameter μ. Expanding (28) permits under H_0 the first order approximation

$$\Lambda_\mu \approx \tilde{l}_\eta{}^T(2\tilde{S}_{x,\mu} - \tilde{S}_{x,\mu}^T \mathrm{Diag}(F_i)\tilde{S}_{x,\mu} - \tilde{P})\tilde{l}_\eta \tag{29}$$

where $\tilde{l}_\eta = \tilde{I}l_\eta = (I - \mathrm{Diag}(F_i)S_{0,\lambda})l_\eta$, see appendix for details. The crucial benefit of the fit $\widehat{\eta}^{(1)}$ in (26) shows in expansion (29) since bias components are cancelling out. The reason is that $\widehat{\eta}^{(0)}$ and $\widehat{\eta}^{(1)}$ have the same smoothing bias under H_0. In addition it should be noted that Λ_μ mainly depends on the bandwidth μ as seen from the components involved in (29). The dependence on λ in turn has minor influence. Finally, it is not difficult to show that Λ_μ is asymptotically normally distributed for $\mu \to 0$. However, as in the previous section, due to slow convergence we prefer drawing inference from the bootstrap version

$$\Lambda_\mu^* = \tilde{l}_\eta^{*T}(2\tilde{S}_{x,\mu} - \tilde{S}_{x,\mu}^T \text{Diag}(F_i)\tilde{S}_{x,\mu} - \tilde{P})\tilde{l}_\eta^*$$

with $l_\eta^* = \tilde{I}_\lambda l_\eta^*$ and l_η^* simulated from the fitted H_0 model with predictor $\widehat{\eta}^{(0)}$.

Power Consideration

We assess the power of the test by considering alternative models of the type H_1 : $E(y|x,u) = h[\gamma_0(u) + Z_x(x)\{\beta_x + n^{-p}\varphi_x(u)\}]$ where $\varphi_x(u)$ is a bounded, smooth but arbitrary function. To ensure identifiability $\varphi_x(u)$ is assumed to have zero mean. We show in the appendix that for $\lambda \to 0$ and $p < 1/2$ the proposed test asymptotically rejects H_0 with probability 1. Hence, as in the previous section we achieve a rate of power which typically holds for parametric tests.

Choice of the Bandwidth

There are two bandwidths involved in this setting. The first, λ, steers the smoothness of the main effect in the semiparametric model. It may be chosen by standard routines like cross validation or the Akaike criterion. The second bandwidth allows for variation of the factorial effects and therefore steers the additional structure of model H_1 compared to H_0. We suggest the Akaike criterion

$$\widehat{\mu} = \arg\max\{\Lambda_\mu - 2E_{H_0}(\Lambda_\mu)\} \tag{30}$$

where $E_{H_0}(\Lambda_\mu) = \text{tr}\{(2\tilde{S}_{x,\mu} - \tilde{S}_{x,\mu}^T \text{Diag}(F_i)\tilde{S}_{x,\mu} - \tilde{P})\text{Diag}(F_i)\}$. To avoid under-smoothing it can be helpful to restrict the range of μ, e.g. by postulating $\widehat{\mu} \geq \widehat{\lambda}$ with $\widehat{\lambda}$ as selected bandwidth for λ. This means the complexity of the factorial varying coefficients $\gamma_x(u)$ is not allowed to exceed the complexity of the main effect $\gamma_0(u)$. A small simulation will supports this setting as well as the use of (30).

3.3 Simulation and Example

Simulation:

We consider the semiparametric logit model $\eta = \gamma_0(u)+x/2$ with main effect $\gamma_0(u) = -0.5+u+\sin(u\pi)/2$. As in the previous section we take u from 30 equidistant points on $[0, 1]$ and at each point of u we simulate y as five repetitions of a binary response for $x = -1$ and $x = 1$. The power of the test is assessed by drawing y from the alternative models $H_{1,a} : \eta = \gamma_0(u) + 0.5 * xu$ and $H_{1,a} : \eta = \gamma_0(u) + x\gamma_0(x)$. Table 4 shows the results based 500 simulations each with 1000 bootstrap replicates. The proposed test shows a powerful behavior and detects non-additive effects of continuous and factorial regressors.

(Table 4)

Example:

We investigate data taken from the German socio economic panel. The binary response y describes whether an unemployed person is reemployed ($y = 1$ for yes). The covariates investigated are the duration of unemployment, u, and the factorial quantities gender, x_1, and nationality, x_2. The focus of interest is to assess whether gender and nationality effects vary with the duration of unemployment. We test the semiparametric model $H_0 : E(y|x_1, x_2, u) = \text{logit}^{-1}\{\gamma_0(u) + x_1\beta_1 + x_2\beta_2\}$ where all effects act additively against the varying coefficient model $H_1 : E(y|x_1, x_2, u) = \text{logit}^{-1}\{\gamma_0(u) + x_1\gamma_1(u) + x_2\gamma_2(u)\}$. We choose $\widehat{\lambda} = 15$ by cross validation and select $\widehat{\mu} = 15$ by (30). This leads to the p-value 0.005. Obviously there is clear evidence that the factorial effects interact with the duration of unemployment. Figure 2 shows the fitted semiparametric model and the corresponding fitted varying coefficient model with bandwidth $\widehat{\mu} = 15$. As seen from the predictors, additivity of the effects of gender, nationality and age may be assumed only for the first 20 months. Afterwards the effect of nationality vanishes and the gender effect decreases. Hence, the factorial effects interact with the continuous covariate so that the semiparametric model seems not adequate for the entire range of duration of unemployment.

(Figure 2)

4 Discussion

We suggest tests for testing parametric or semiparametric models with continuous and factorial regressors against smooth alternatives. We fit the alternative model by locally fitting the H_0 model. In both settings this allows to neglect the smoothing bias in general. The objective of this fit is on testing and one should keep in mind that the fit of the H_1 model is not necessarily a good fit when the objective is estimation solely. This particularly holds since the bandwidth selection criteria (17) and (30) used in the paper emphasize the testing problem, i.e. the difference between the parametric and nonparametric fit, while the bias-variance trade off is a minor issue here.

A Technical Details

A.1 Appendix for §2

Derivation of Expansion (10)

We have

$$\begin{aligned} 0 &= \sum_j \omega_{\lambda,ij} W_j^T l_{\eta,j}(W_j \widehat{\beta}_i) \approx \sum_j \omega_{\lambda,ij} W_j^T \{l_{\eta,j} - F_j(Z_j V_j \widehat{\beta}_i - Z_j \gamma_j)\} \\ \Leftrightarrow \widehat{\beta}_i &\approx \mathbf{F}_{(i)}^{-1} \left(\sum_j \omega_{\lambda,ij} W_j^T l_\eta + \sum_j \omega_{\lambda,ij} W_j^T F_j Z_j \gamma_j \right) \end{aligned} \tag{31}$$

where $\mathbf{F}_{(i)} = \sum_j \omega_{\lambda,ij} W_j^T F_j W_j$ and $l_{\eta,j} = l_{\eta,j}(\eta_j) = l_{\eta,j}(Z_j\gamma_j)$. Since $V(u)$ is supposed to have a row diagonal structure, i.e. in each column there is only a single non-zero element and 1 is element of each row, one gets $V_i \mathbf{F}_{(i)}^{-1} \sum_j \omega_{\lambda,ij} W_j^T F_j Z_j = I$ with I as identity matrix. This in turn permits to write $\gamma_i = V_i \mathbf{F}_{(i)}^{-1} \sum_j \omega_{\lambda,ij} W_j^T F_j Z_j \gamma_i$ which proves (10) with (31).

Moments of the Likelihood Ratio Statistics

Formula (14) gives the first order approximation $\Lambda_\lambda \approx l_\eta^T M l_\eta$ where $M = (2S_\lambda - S_\lambda \mathrm{Diag} S_\lambda^T - P)$. Matrix P is a projection type matrix, i.e. we have $P\mathrm{Diag}(F_i)P = P$ or $S_\lambda \mathrm{Diag}(F_i)P = P$. Let the elements of M be denoted by M_{ij} and set $\tilde{M} = M\mathrm{Diag}(F_i)$. Derivation of the expectation of Λ_λ is direct since $E(l_\eta^T M l_\eta) = \mathrm{tr}(\tilde{M})$. The second order moment of Λ_λ equals $E(\Lambda_\lambda^2) \approx \mathrm{tr}(\tilde{M})^2 + \mathrm{tr}(\tilde{M}\tilde{M}) + \mathrm{tr}(\tilde{M}\tilde{M}^T)$ which yields the variance $\mathrm{Var}(\Lambda_\lambda) \approx \mathrm{tr}(\tilde{S}_\lambda \tilde{S}_\lambda + \tilde{S}_\lambda \tilde{S}^T{}_\lambda) - 2q$. In the same fashion one gets higher order cumulants, e.g. the third cumulant $\mathrm{Cum}_3(\Lambda_\lambda) \approx \mathrm{tr}(2\tilde{S}_\lambda \tilde{S}_\lambda \tilde{S}_\lambda + 6\tilde{S}_\lambda \tilde{S}_\lambda \tilde{S}_\lambda^T) - 8q$. The definition of S_λ given in (11) shows that the diagonal elements of S_λ have order $O\{(n\lambda)^{-1}\}$, neglecting boundary points, so that $\mathrm{Cum}_3(\Lambda_\lambda) = O(\lambda^{-1})$. Moreover, assuming F_i to be bounded away from zero one also has $\mathrm{Var}(\Lambda_\lambda)^{-3/2} =$

$O(\lambda^{3/2})$ which in turn proves the asymptotic normality stated in (15).

Power Consideration

Let us assume that H_1 holds which implies that the bias B_λ in (12) does not vanish. Moreover, the estimates in the H_0 model fulfill $\widehat{\eta}^{(0)} - \eta = Pl_\eta + B_\infty$ where $B_\infty = \lim_{\lambda \to \infty} B_\lambda = P\eta - \eta$ is the bias which occurs under H_1 when fitting the H_0 model. The likelihood ratio now equals

$$\begin{aligned} \Lambda_\lambda &\approx l_\eta^T M l_\eta + 2 l_\eta^T \{I - S_\lambda \mathrm{Diag}(F_i)\} B_\lambda - 2 l_\eta^T \{I - P_\beta \mathrm{Diag}(F_i)\} B_\infty \\ &\quad - B_\lambda^T \mathrm{Diag}(F_i) B_\lambda + B_\infty^T \mathrm{Diag}(F_i) B_\infty \\ &= l_\eta^T M l_\eta + bias^2 + O_p(n^{1/2}\lambda^4) \end{aligned}$$

with $bias^2 = B_\infty^T \mathrm{Diag}(F_i) B_\infty - B_\lambda^T \mathrm{Diag}(F_i) B_\lambda$ and M as defined above. The latter simplification above holds since $\{I - P\mathrm{Diag}(F_i)\}B_\infty = 0$ and $\{I - S_\lambda \mathrm{Diag}(F_i)\}B_\lambda = O(\lambda^2)O(B_\lambda)$. Consider now the alternative model with $\gamma(u) = V(u)\beta + \varphi(u)n^{-p}$ with $p < 1/2$ and $0 < a \leq \sum_j^n |\varphi_j|/n \leq b < \infty$ where $\varphi_j = \varphi(u_j)$. For identifiability reasons we also assume that $V(u)\beta$ and $\varphi(u)$ are orthogonal in the sense $\sum_j^n W_j^T F_j Z_j \varphi_j = 0$. Reflecting the definition of $\gamma(u)$ and following standard kernel smoothing arguments one gets the asymptotic order $B_\lambda^T \mathrm{Diag}(F_i) B_\lambda = O(\lambda^4 n^{1-2p})$ and $B_\infty^T \mathrm{Diag}(F_i) B_\infty = O(n^{1-2p})$. Since $E_{H_0}(\Lambda_\lambda) = O(\lambda^{-1})$ we select from (17) an optimal bandwidth λ with order $\lambda = O(n^{(-1+2p)/5})$. This in turn gives $bias^2$ of order $O(n^{1-2p})$. Reflecting now that the variance of Λ_λ coincides under H_0 and H_1 with asymptotic order $O(\lambda^{-1})$ provides $bias^2/\sqrt{\mathrm{Var}(\Lambda_\lambda)} = O(n^{(11-22p)/10})$. This shift tends to infinity for $p < 1/2$ so that with

$$P\left(\frac{\Lambda_\lambda - E_{H_0}(\Lambda_\lambda)}{\sqrt{\mathrm{Var}(\Lambda_\lambda)}} \leq z\right) = P\left(\frac{\Lambda_\lambda - E_{H_1}(\Lambda_\lambda)}{\sqrt{\mathrm{Var}(\Lambda_\lambda)}} \leq z - \frac{bias^2}{\sqrt{\mathrm{Var}(\Lambda_\lambda)}}\right) \qquad (32)$$

the corresponding test rejects H_0 asymptotically with probability one.

A.2 Appendix for §3

Derivation of Formula (22) and (24)

For asymptotic considerations it is helpful to incorporate the dependence of $\widehat{\gamma}_{0,i}$ on β_x in the notation. Let $\widehat{\gamma}_{0,i|\beta_x}$ be the solution of (20) for fixed β_x. Differentiating (20) gives in first order approximation,

$$0 \approx \sum_j \omega_{\lambda,ij} V_{0,j}^T F_j \left(V_{0,j} \frac{\partial \widehat{\beta}_{0,i|\beta_x}}{\partial \beta_x^T} + Z_{x,j} \right) \tag{33}$$

with $\widehat{\beta}_{0,i|\beta_x}$ as solution of (20) and $\widehat{\gamma}_{0,i|\beta_x} = V_{0,i}\widehat{\beta}_{0,i|\beta_x}$. Solving (33) for $\partial\widehat{\beta}_{0,i|\beta_x}/\partial\beta_x^T$ provides (22). With $\widehat{\gamma}_{0,i} := \widehat{\gamma}_{0,i|\widehat{\beta}_x}$ we denote the final estimate for $\gamma_{0,i}$ where $\widehat{\beta}_x$ solves (21). By expansion we get in first order approximation

$$\widehat{\gamma}_{0,i|\widehat{\beta}_x} \approx \widehat{\gamma}_{0,i|\beta_x} + \frac{\partial \widehat{\gamma}_{0,i|\beta_x}}{\partial \beta_x^T}(\widehat{\beta}_x - \beta_x) + o_p(\widehat{\beta}_x - \beta_x)$$

where β_x denotes the true parameter here. Making use of the definition $\tilde{Z}_{x,i} = Z_{x,i} + (\partial\widehat{\gamma}_{0,i|\beta_x})/(\partial\beta_x^T)$ one obtains for $\widehat{\eta}_i^{(0)} = \widehat{\gamma}_{0,i} + Z_{x,i}\widehat{\beta}_x$ the approximation

$$\widehat{\eta}_i^{(0)} - \eta_i \approx \widehat{\gamma}_{0,i|\beta_x} - \gamma_{0,i} + \tilde{Z}_{x,i}(\widehat{\beta}_x - \beta_x). \tag{34}$$

We first expand (20) by taking β_x as given true parameter. This gives in first order approximation

$$\widehat{\gamma}_{0,i|\beta_x} \approx \gamma_{0,i} + V_{0,i}\mathbf{F}_{00(i)}^{-1} \left\{ \sum_j \omega_{\lambda,ij} V_{0,j}^T l_{\eta,j} \right\} + b_{0(i)} \tag{35}$$

where $\mathbf{F}_{00(i)} = \sum_j \omega_{\lambda,ij} V_{0,j}^T F_j V_{0,j}$ and $b_{0,(i)} = V_{0,i}\mathbf{F}_{00(i)}^{-1} \sum_j \omega_{\lambda,ij} V_{0,j}^T F_j (\gamma_{0,j} - \gamma_{0,i})$ occurs as smoothing bias. With $S_{0,\lambda}$ denoting the generalized smoothing matrix as given in (23) one may rewrite (35) in matrix form $\widehat{\gamma}_{0|\beta_x} - \gamma_0 \approx S_{0,\lambda} l_\eta + B_{0,\lambda}$ where $\gamma_0 = (\gamma_{0,1}, \ldots, \gamma_{0,n})^T$, $\widehat{\gamma}_{0|\beta_x} = (\widehat{\gamma}_{0,1|\beta_x}, \ldots, \widehat{\gamma}_{0,n|\beta_x})^T$ and $B_{0,\lambda} = S_{0,\lambda}\mathrm{Diag}(F_i)\gamma_0 - \gamma_0$. The next step is to expand of the profile estimation function (21) which gives in first order approximation

$$\begin{aligned}
\widehat{\beta}_x - \beta_x &\approx \left\{\sum_i \widetilde{Z}_{x,i}^T F_i \widetilde{Z}_{x,i}\right\}^{-1} \left\{\sum_i \widetilde{Z}_{x,i}^T l_{\eta,i}(\widehat{\gamma}_{0,i|\beta_x} + Z_{x,i}\beta_x)\right\} \\
&\approx \left\{\sum_i \widetilde{Z}_{x,i}^T F_i \widetilde{Z}_{x,i}\right\}^{-1} \left[\sum_i \{\widetilde{Z}_{x,i}^T l_{\eta,i} - \widetilde{Z}_{x,i}^T F_i(\widehat{\gamma}_{0,i|\beta_x} - \gamma_{0,i})\}\right] \qquad (36) \\
&\approx \left\{\sum_i \widetilde{Z}_{x,i}^t F_i \widetilde{Z}_{1,i}\right\}^{-1} \left(\sum_i \widetilde{Z}_{x,i}^T \tilde{l}_{\eta,i}\right) \qquad (37)
\end{aligned}$$

where $\tilde{l}_{\eta,i} = l_{\eta,i} - F_i S_{0,\lambda;i\bullet} l_\eta$ with $S_{0,\lambda;i\bullet}$ denoting the ith row of $S_{0,\lambda}$. The simplification from (36) to (37) holds up to the considered asymptotic order, as generally proven in Severini & Wong (1992). This implies in particular, that the effect of smoothing on the parametric fit $\widehat{\beta}$ is only of second order, i.e. $E(\widehat{\beta} - \beta) = O(\lambda^4) + O(n^{-1})$. In particular (37) follows since the component

$$\begin{aligned}
\sum_i \tilde{Z}_{x,i}^T F_i b_{0(i)} &= \tilde{Z}_x^T \mathrm{Diag}(F_i)\{S_{0,\lambda}\mathrm{Diag}(F_i)\gamma_0 - \gamma_0\} \\
&= -Z_x^T\{I - S_{0,\lambda}\mathrm{Diag}(F_i)\}^T \mathrm{Diag}(F_i)(I - S_{0,\lambda})\mathrm{Diag}(F_i)\gamma_0 \\
&= O(n\lambda^4) \qquad (38)
\end{aligned}$$

is of negligible asymptotic order, where we made use of the property $\sum_i \omega_{\lambda,ij} = \sum_i \omega_{\lambda,ji}\{1 + O(\lambda^2)\}$. Inserting(35) and (37) in (34) finally proves (24).

Properties of estimate $\widehat{\eta}^{(1)}$ defined in (26)

Let $\tilde{I}_\lambda = I - \mathrm{Diag}(F_i)S_{0,\lambda}$ and define $S = S_{0,\lambda} + \tilde{S}_{x,\mu}\tilde{I}_\lambda$. Then in first order approximation

$$\begin{aligned}
\widehat{\eta}^{(1)} &= \widehat{\eta}^{(0)} + S\widehat{l}_\eta^{(0)} + S\mathrm{Diag}(F_i)\widehat{\eta}^{(0)} - \widehat{\eta}^{(0)} \\
&\approx \eta + S\{l_\eta - \mathrm{Diag}(F_i)(\widehat{\eta}^{(0)} - \eta)\} + S\mathrm{Diag}(F_i)\widehat{\eta}^{(0)} - \eta \\
&\approx \eta + Sl_\eta + S\mathrm{Diag}(F_i)\eta - \eta \qquad (39)
\end{aligned}$$

Under H_0 we have $\eta = \gamma_0 + Z_x\beta$ which allows to simplify the bias component in (39). Making use of $S_{x,\mu}\tilde{I}_\lambda \text{Diag}(F_i)Z_x = S_{x,\mu}\text{Diag}(F_i)\tilde{Z}_x = \tilde{Z}_x$ provides

$$S\text{Diag}(F_i)\eta - \eta \;=\; B_{0,\lambda} + \tilde{S}_{x,\mu}\text{Diag}(F_i)B_{0,\lambda} \tag{40}$$

with $B_{0,\lambda} = S_{0,\lambda}\text{Diag}(F_i)\gamma_0 - \gamma_0$. We show now that for $\lambda \to 0$ the component $\tilde{S}_{x,\mu}\text{Diag}(F_i)B_{0,\lambda}$ in (40) has negligible asymptotic order which in turn proves (27). Assuming $\mu \to \infty$ one gets from (38) $\tilde{S}_{x,\infty}\text{Diag}(F_i)B_{0,\lambda} = \tilde{P}\text{Diag}(F_i)B_{0,\lambda} = O(\lambda^4)$. Let on the other hand $\mu \to 0$. We define with $E(Z_x|u_i)$ the mean of the covariates given $u = u_i$, i.e. we have $S_{0,\lambda}\text{Diag}(F_i)Z_x = E(Z_x)\{1 + O(\lambda^2)\}$ where $E(Z_x)^T = \{E(Z_x|u_1)^T, \ldots, E(Z_x|u_n)^T\}$. Moreover we extract the dominating components of the bias $B_{0,\lambda}$ by writing $B_{0,\lambda} = \delta\lambda^2 + O(\lambda^4)$ where $\delta = (\delta_1, \ldots \delta_n)^T$ with $\delta = O(1)$ and boundary effects are neglected. For instance for linear smoothing one has $\delta_i = \gamma_0''(u_i)/2$. This notation permits

$$\begin{aligned}\frac{1}{n\mu}\sum_j \omega_{\mu,ij}\tilde{Z}_{x,j}F_j b_{0,(j)} &= \frac{\lambda^2}{n\mu}\sum_j \omega_{\mu,ij}\{Z_{x,j} - E(Z_x|u_j)\}F_j\delta_j + O(\lambda^4) \\ &= \lambda^2 O(\mu^2) + O(\lambda^4)\end{aligned} \tag{41}$$

so that $\tilde{S}_{x,\mu}\text{Diag}(F_i)B_{0,\lambda} = O(\lambda^2)\{O(\mu^2) + O(\lambda^2)\}$. Hence the latter component in (40) is of negligible order for both, $\mu \to \infty$ and $\mu \to 0$, which proves (27)

Expansion of the Likelihood ratio (29)

Expansion of the likelihood ratio (28) permits

$$\begin{aligned}\Lambda_\mu &\approx 2l_\eta^T(\widehat{\eta}^{(1)} - \widehat{\eta}^{(0)}) \\ &\quad -(\widehat{\eta}^{(1)} - \eta)^T\text{Diag}(F_i)(\widehat{\eta}^{(1)} - \eta) + (\widehat{\eta}^{(0)} - \eta)^T\text{Diag}(F_i)(\widehat{\eta}^{(0)} - \eta) \qquad (42) \\ &\approx 2l_\eta^T(\tilde{S}_{x,\mu} - \tilde{P})\tilde{l}_\eta - \{l_\eta^T(S_{0,\lambda} + \tilde{S}_{x,\mu}\tilde{I}_\lambda)^T + B_0\}^T\text{Diag}(F_i)\{(S_{0,\lambda} + \tilde{S}_{x,\mu}\tilde{I}_\lambda)l_\eta + B_0\} \\ &\quad +\{l_\eta^T(S_{0,\lambda} + \tilde{P}\tilde{I}_\lambda)^T + B_0\}^T\text{Diag}(F_i)\{(S_{0,\lambda} + \tilde{P}\tilde{I}_\lambda)l_\eta + B_0\}\end{aligned}$$

$$= \quad \tilde{l}_\eta^T (2\tilde{S}_{x,\mu} - \tilde{S}_{x,\mu}^T \mathrm{Diag}(F_i)\tilde{S}_{x,\mu} - \tilde{P})\tilde{l}_\eta + B_0^T \mathrm{Diag}(F_i)(\tilde{P} - \tilde{S}_{x,\mu})\tilde{l}_\eta \tag{43}$$

with $\tilde{l}_\eta = \tilde{I}_\lambda l_\eta$. The squared bias components above cancel out since under H_0 both estimates $\widehat{\eta}^{(0)}$ and $\widehat{\eta}^{(1)}$ have the same smoothing bias. Moreover, the later component in (43) has zero expectation and is of negligible asymptotic order. This yields (29) as first order approximation for the likelihood ratio.

Power Consideration

Let model H_1 hold with $\eta_i = \gamma_{0,i} + Z_{x,i}(\beta_x + n^{-p}\varphi_{x,i})$, where $\varphi_{x,i} = \varphi_x(u_i)$ fulfills the orthogonality condition $\sum_i \tilde{Z}_{x,i}^T F_i \tilde{Z}_{x,i}\varphi_i = 0$ to ensure identifiability. We assume $\lambda \to 0$ and $\mu \to 0$ and take $\beta_x = 0$ for simplicity. With η_x we define the vector $n^{-p}(Z_{x,1}\varphi_{x,1}, \ldots, Z_{x,n}\varphi_{x,n})^T$. Using results from above we can write $\widehat{\eta}^{(1)}$ under H_1 as

$$\widehat{\eta}^{(1)} \approx \eta + Sl_\eta + S\mathrm{Diag}(F_i)\eta - \eta \quad \approx \quad \eta + Sl_\eta + B_{0,\lambda} + \tilde{B}_{x,\mu},$$

with bias components $B_{0,\lambda} = S_{0,\lambda}\mathrm{Diag}(F_i)\gamma_0 - \gamma_0$ and $\tilde{B}_{x,\mu} = \tilde{S}_{x,\mu}\mathrm{Diag}(F_i)\tilde{\eta}_x - \tilde{\eta}_x$ where $\tilde{\eta}_x = \{I - S_{0,\lambda}\mathrm{Diag}(F_i)\}\eta_x$. In the same fashion we get $\widehat{\eta}^{(0)} \approx (S_{0,\lambda} + \tilde{P}\tilde{I}_\lambda)l_\eta + B_{0,\lambda} + \tilde{B}_{x,\infty}$ with $\tilde{B}_{x,\infty} = \lim_{\mu\to\infty}\tilde{B}_{x,\mu} = \tilde{P}\mathrm{Diag}(F_i)\tilde{\eta}_x - \tilde{\eta}_x$. Standard kernel smoothing arguments allow to derive $\tilde{\eta}_{x,i} = n^{-p}\tilde{Z}_{x,i}\varphi_{x,i}\{1 + O(\lambda^2)\}$ and $\tilde{S}_{x,\mu}\mathrm{Diag}(F_i)\tilde{\eta}_x = \{\tilde{\eta}_x + O(\mu^2)\}\{1 + O(\lambda^2)\}$, where boundary effects are neglected. This implies the asymptotic order $\tilde{B}_{x,\mu} = n^p\{O(\mu^2) + O(\lambda^2)\}$ and $\tilde{B}_{x,\infty} = n^p\{1 + O(\lambda^2)\}$. Making use of (38) we find the bias component $B_{0,\lambda}$ to be orthogonal to $\tilde{Z}_x$ in the sense $n^{-1}\tilde{Z}_x^T\mathrm{Diag}(F_i)B_{0,\lambda} = O(\lambda^4)$. Moreover, reflecting (41) one gets $\tilde{S}_{x,\mu}\mathrm{Diag}(F_i)B_{0,\lambda} = O(\lambda^2)\{O(\mu^2) + O(\lambda^2)\}$ which in turn provides that the bias component $B_{0,\lambda}$ is orthogonal to $\tilde{B}_{x,\mu}$, i.e. we have $n^{-1}\tilde{B}_{x,\mu}^T\mathrm{Diag}(F_i)B_{0,\lambda} = n^{-p}O(\lambda^2)\{O(\lambda^2) + O(\mu^2)\}^2$. This orthogonality also holds for $\tilde{B}_{x,\mu}$ substituted by $\tilde{B}_{x,\infty}$. Inserting $\widehat{\eta}^{(1)}$ and $\widehat{\eta}^{(0)}$

in (42) shows now by making use of the above orthogonalities

$$E_{H_1}(\Lambda_\lambda) \approx E_{H_0}(\Lambda_\lambda) + bias^2$$

where $bias^2 = \tilde{B}^T_{x,\infty}\mathrm{Diag}(F_i)\tilde{B}_{x,\infty} - \tilde{B}_{x,\mu}\mathrm{Diag}(F_i)\tilde{B}_{x,\mu}$. The bias components have the asymptotic order $\tilde{B}_{x,\mu}\mathrm{Diag}(F_i)\tilde{B}_{x,\mu} = O(n^{1-2p})\{O(\mu^2) + O(\lambda^2)\}^2$ and $\tilde{B}^T_{x,\infty}\mathrm{Diag}(F_i)\tilde{B}_{x,\infty} = O(n^{1-2p})\{1 + O(\lambda^2)\}^2$ and $E_{H_0}(\Lambda_\lambda) = O(\mu^{-1})$. Since $\lambda \to 0$ is assumed we can neglect the effect of λ in the sequel. The optimal bandwidth μ chosen by (30) has order $\mu = O(n^{(-1+2p)/5})$ which implies that $bias^2 = O(n^{(11-22p)/10})$. As in Section 2 we get with (32) that the test rejects H_0 asymptotically with probability one, as stated.

References

Aerts, M., Claeskens, G., & Hart, J. D. (1998). Testing the fit of a parametric function. *J. Amer. Statist. Assoc.*, (to appear).

Azzalini, A. & Bowman, A. W. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. B 55*, 549–557.

Azzalini, A., Bowman, A. W., & Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika 76*, 1–11.

Barndorff-Nielsen, O. E. & Cox, D. R. (1989). *Asymptotic Techniques for use in Statistics.* Chapman & Hall.

Bowman, A. & Young, S. (1996). Graphical comparison of nonparametric curves. *Appl. Statist. 45*, 83–98.

Bowman, A. W. & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the kernel approach with S-Plus illustrations.* Oxford University Press.

Carroll, R. J., Ruppert, D., & Welsh, A. H. (1988). Local estimating equations. *J. Amer. Statist. Assoc. 93*, 214–227.

le Cessie, S. & van Houwelingen, J. (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics 47*, 1267–1282.

Cuzick, J. (1992). Semiparametric additive regression. *J. Roy. Statist. Soc. B 54*, 831–843.

Dette, H. & Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *J. Roy. Statist. Soc. B 60*, 693–708.

Eubank, R. L., Hart, J., & LaRiccia, V. (1993). Testing goodness of fit via nonparametric function estimation techniques. *Comm. Statist. 22*, 3327–3354.

Eubank, R. L. & Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc. 85*, 387–392.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London: Chapman & Hall.

Fan, J., Heckman, N. E., & Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc. 90*(429), 141–150.

Fan, Y. & Li, Q. (1996). Consistent model specification test: Omitted variables and semiparametric functional forms. *Econometrica 64*, 865–90.

Firth, D., Glosup, J., & Hinkley, D. (1991). Model checking with nonparametric curves. *Biometrika 78*(2), 245–52.

Härdle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Stat. 21*, 1926–1947.

Härdle, W., Mammen, E., & Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *J. Amer. Statist. Assoc. 93*, 1461–1474.

Hart, D. J. & Wehrly, T. E. (1992). Kernel regression when the boundary region is large, with an application to testing the adequacy of polynomial models. *J. Amer. Statist. Assoc. 87*, 1018–1024.

Hart, J. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer Verlag.

Hastie, T. & Tibshirani, R. (1990). *Generalized Additive Models.* London: Chapman and Hall.

Hastie, T. & Tibshirani, R. (1993). Varying–coefficient models. *J. Roy. Statist. Soc. B 55*, 757–796.

Heckman, N. (1986). Spline smoothing in partly linear models. *J. Roy. Statist. Soc. 48*, 244–248.

Hunsberger, S. (1995). Semiparametric regression in likelihood based models. *J. Amer. Statist. Assoc. 89*, 1354–1365.

Kauermann, G. & Tutz, G. (1998a). Local likelihood estimation in varying-coefficient models including additive bias correction. *J. of Nonparametric Stat. (to appear).*

Kauermann, G. & Tutz, G. (1998b). On model diagnostics using varying coefficient models. *Biometrika 86*, 119–128.

Müller, H.-G. (1992). Goodnes-of-fit diagnostics for regression models. *Scand. J. Statist. 19*, 157–172.

Raz, J. (1990). Testing for no effect when estimating a smooth function by non-parametric regression: A randomization approach. *J. Amer. Statist. Assoc. 85*, 132–138.

Severini, T. A. & Staniswalis, J. G. (1994). Quasi–likelihood estimation in semi-parametric models. *J. Amer. Statist. Assoc. 89*, 501–511.

Severini, T. A. & Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist. 20*, 1768–1802.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. B 50*, 413–436.

Staniswalis, J. G. & Severini, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc. 86*, 684–692.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist. 25*, 613–641.

	smooth test		parametric test	
Model	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 10\%$
H_0	6.3	12.1	4.7	10.0
$H_{1,a}$	18.7	29.7	23.7	33.0
$H_{1,b}$	20.0	31.7	5.2	10.7
$H_{1,c}$	23.0	33.7	7.5	16.5

Table 1: Probability of rejection in a simulation study for testing $H_0 : \eta = \beta_0 + u\beta_u + x\beta_x$

Model	$\alpha = 5\%$	$\alpha = 10\%$
H_0	5.2	10.2
$H_{1,a}$	23.7	33.4
$H_{1,b}$	22.8	35.2

Table 2: Probability of rejection in a simulation study for testing $H_0 : \eta = \gamma_0(u) + x\beta_x$

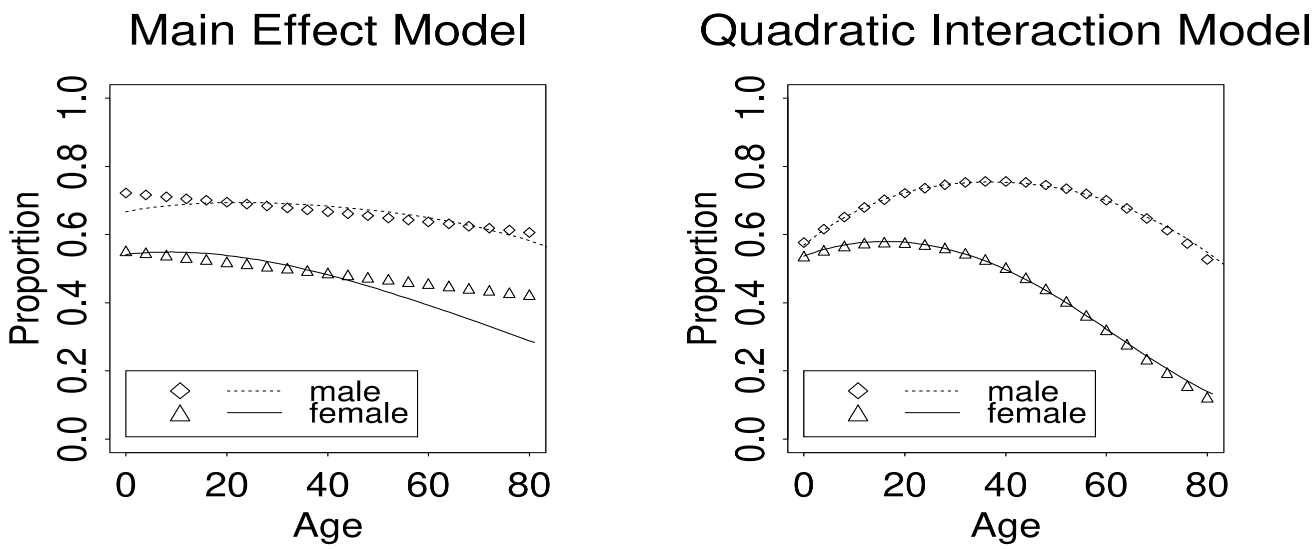

Figure 1: Fitted probability of human parasitic worm infestation. Lines correspond to the fitted H_1 model, points ($\diamond$, $\triangle$) represent the fitted H_0 model.

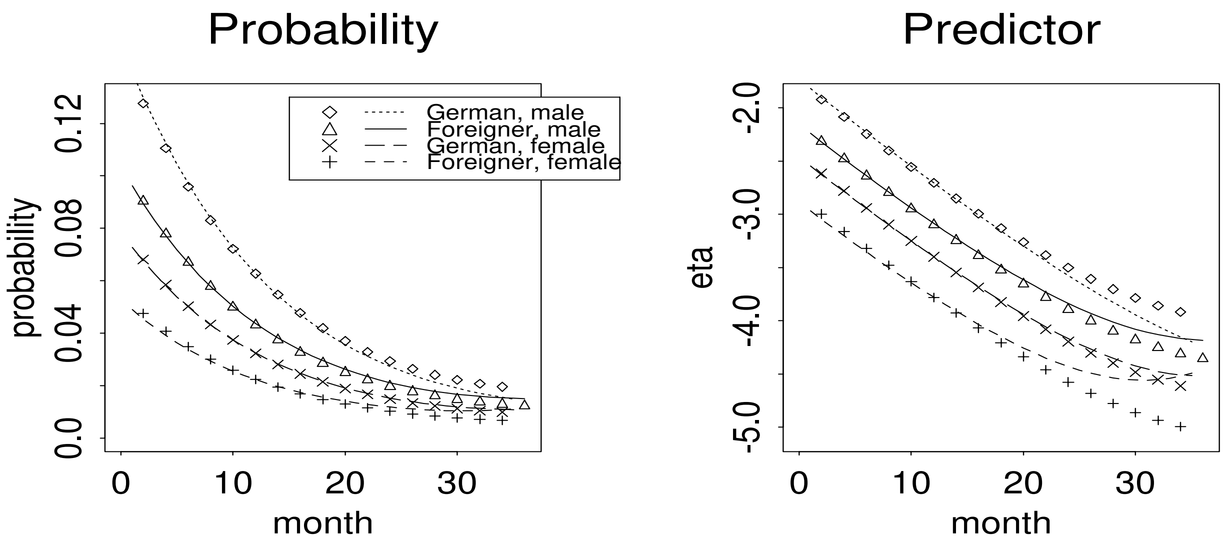

Figure 2: Estimated probabilities and fitted predictors for unemployment data. Points ($\diamond$, $\triangle$, $\times$, $+$) show the semiparametric fit and lines represent the locally fitted semiparametric model.