

SFB 649 Discussion Paper 2013-028

Analysis of Deviance in Generalized Partial Linear Models

Wolfgang Karl Härdle*
Li-Shan Huang**



* Humboldt-Universität zu Berlin, Germany
** National Tsing Hua University, Taiwan

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Analysis of Deviance in Generalized Partial Linear Models

WOLFGANG KARL HÄRDLE

Center for Applied Statistics & Economics, Humboldt University, Berlin, GERMANY

and School of Business, Singapore Management University, SINGAPORE

haerdle@wiwi.hu-berlin.de

AND

LI-SHAN HUANG¹

Institute of Statistics, National Tsing Hua University, TAIWAN

Lhuang@stat.nthu.edu.tw

May 20, 2013

Abstract

We develop analysis of deviance tools for generalized partial linear models based on local polynomial fitting. Assuming a canonical link, we propose expressions for both local and global analysis of deviance, which admit an additivity property that reduces to ANOVA decompositions in the Gaussian case. Chi-square tests based on integrated likelihood functions are proposed to formally test whether the nonparametric term is significant. Simulation results are shown to illustrate the proposed chi-square tests. The methodology is applied to German Bundesbank Federal Reserve data.

Keywords: ANOVA decomposition, Integrated likelihood, Link function, Local polynomial

AMS 2000 subject classifications: Primary 62G08; secondary 62J12

JEL classification: C00, C14, C50, C58

ACKNOWLEDGMENT: The work was conceived during the visit of the first author to the Humboldt University supported by CRC 649 “Economic Risk.” The support is greatly appreciated. The first author was partially supported by National Science Council (NSC 101-2118-M-007-002-MY2) in Taiwan and both authors were partially supported by CRC 649 “Economic Risk.”

¹Corresponding author, address: No.101 Kuang-Fu Road, Section 2, Hsinchu City, TAIWAN 30013.

1 Introduction

Generalized linear models (GLMs) (McCullagh and Nelder, 1989) are a large class of statistical models for relating a response variable to linear combinations of predictor variables. The models allow the data to follow some probability distributions other than the Gaussian distribution in linear models, such as the Binomial and Poisson, etc. For example, logistic regression under the Binomial assumption is a subclass of GLMs and is commonly used to model the probability of events given a linear function of the explanatory variables. An example that we address in this paper is the probability of default (PD) for corporate bankruptcy prediction, known as rating or scoring, to estimate the financial situation of a company and the probability that a company defaults on its obligations within a certain period. A major disadvantage of this approach is that the prespecified linear form of predictors is too restrictive. Companies that grow too fast or too slow tend to have a higher probability of default and hence models allowing for nonlinear effects may be more appropriate.

Generalized partial linear models (GPLMs) (Green and Silverman, 1994; Carroll et al., 1997; Härdle et al., 2004) allow for a nonparametric component for a continuous covariate while retaining the ease of linear relationships for the remaining variables. It is more flexible than parametric GLMs and simpler than generalized additive models (GAMs) (Hastie and Tibshirani, 1990) which include multiple nonparametric components. In applying GPLMs to, e.g. PD data, inference tools to examine whether the nonparametric term is significant are of interest. Though Hastie and Tibshirani (1990) briefly discussed Analysis of Deviance (ANODEV) for GPLMs, they note that “the distribution theory, however, is undeveloped” and “informal deviance tests with some heuristic justification” are adopted. Härdle et al. (1998) introduced a test statistic for testing linearity in GPLMs with an asymptotically normal distribution. Li and Liang (2008) address assessing significance of the nonparametric term for the GPLMs in the local polynomial settings by extending generalized likelihood ratio tests (Fan et al., 2001), which have asymptotic chi-square distributions. We note that testing in the GPLMs is relatively less developed as compared to the special case of partial linear models (PLMs) under Gaussian distributions (Härdle et al., 2004). Hence, there is a need for developing inference tools for applications of GPLMs, such as the PD data.

Based on the local polynomial approach (Fan and Gijbels, 1996) and assuming a canonical

link in GPLMs, we propose local and global expressions for ANODev, with the latter obtained by integrating the corresponding local likelihood quantities. This mimics the “integrated likelihood” approach discussed by Severini (2007) and Lehmann (2006). Though the idea of local likelihood for GLMs has been around for some time (Hastie and Tibshirani, 1987; Loader, 1999), to our knowledge, we are not aware of using the integrated likelihood approach to combine the information of local likelihood in the smoothing literature. In the case of the Gaussian distribution with an identity link, the proposed local and global ANODev becomes local and global analysis of variance (ANOVA) (Huang and Davidson, 2010) for PLMs and F -tests were justified between nested models, similar to classical ANOVA in linear models. Based on the proposed global ANODev, integrated likelihood ratio tests with asymptotic chi-square distributions are proposed for checking whether the nonparametric term is significant. Our work extends the classical ANODev for GLMs to GPLMs with theoretical justifications.

The organization of this paper is as follows. In Section 2, utilizing the ANODev for nested hypotheses in parametric GLMs by Simon (1973), we propose local and global ANODev for nonparametric GLMs in Theorem 1 for the simpler case with the nonparametric term as the only predictor. By combining local likelihood through integration, as a by-product, new estimators for the canonical parameter and response mean are given in equation (11), and Theorem 2 shows that the integrated likelihood quantities are asymptotically global likelihood quantities with the new estimator. Theorem 3 proposes a chi-square test based on integrated likelihood functions for testing whether the nonparametric term is significant. Section 3 presents an extension of Theorems 1-3 to Theorems 4 and 5 for GPLMs. In Section 4, we illustrate the potential usefulness of the new tests with simulation data and Section 5 applies the methodology to 2002 German Bundesbank Federal Reserve data.

2 Nonparametric Analysis of Deviance

Let $(x_1, y_1), \dots, (x_n, y_n)$ be independent data pairs with the conditional density of Y given $X = x$ from a one-parameter exponential family:

$$L(y; \theta) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}, \quad (1)$$

where θ is the canonical parameter, $a(\cdot) > 0$, $b(\cdot)$ and $c(\cdot)$ are known functions, and ϕ is known or a nuisance parameter. We first briefly review parametric GLMs from McCullagh and Nelder (1989), in which a transformation $G(\cdot)$ of $\mu(x) = \mathbb{E}(Y|X = x)$ is modeled parametrically via, e.g., a linear function, $G\{\mu(x)\} = b_0 + b_1x$, and $G(\cdot)$ is called the “link” function. If $G(\cdot) = (b')^{-1}(\cdot)$, then G is the canonical link function that links θ to the linear predictor since $b'\{\theta(\cdot)\} = \mu(\cdot)$. Without loss of generality, we assume that G is the canonical link function throughout the paper.

Let $\ell(y, \theta) = \log L(y, \theta)$, $\hat{\theta}$ denote the fitted value of θ with corresponding \hat{y} , and $\tilde{\theta}$ when the fitted value equals to the observed y . The deviance D for a fitted parametric GLM is defined as

$$D(\mathbf{y}, \hat{\mathbf{y}}) = 2 \sum_i \{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\}, \quad (2)$$

where $\mathbf{y} = (y_1, \dots, y_n)^\top$ and $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)^\top$. In the Gaussian case, G is the identity link and $D = \sum_i (y_i - \hat{y}_i)^2$. Hence the deviance quantity (2) in GLMs is analogous to the sum of squares (SS) of residuals in linear models. Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^\top$. Let us now turn to testing hypotheses about $\boldsymbol{\theta}$. Assume that $D_1 = \inf_{\boldsymbol{\theta} \in \Omega_1} D$ and D_2 the deviance for $\boldsymbol{\theta} \in \Omega_2$ with $\Omega_2 \subseteq \Omega_1$. In GLMs, ANODev usually refers to comparing two nested models and inference may be based on the difference $D_2 - D_1$, which is simply the log likelihood ratio statistic with an asymptotic χ^2 distribution. In the Gaussian case, ANODev becomes a comparison of two residual SS. Note that the conventional ANODev is generally not parallel to ANOVA in linear models, in the sense that ANODev does not have a complete set of SS quantities (SS total = SS regression + SS residuals).

An attempt to mimic ANOVA can be based on the Kullback-Leibler (KL) divergence of two probability distributions with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$:

$$KL(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = 2\mathbb{E}_{\boldsymbol{\mu}_1} [\ell\{\mathbf{y}, G^{-1}(\boldsymbol{\mu}_1)\} - \ell\{\mathbf{y}, G^{-1}(\boldsymbol{\mu}_2)\}], \quad (3)$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are treated as fixed values and $\mathbb{E}_{\boldsymbol{\mu}_1}$ is expectation with respect to \mathbf{y} when $\boldsymbol{\mu} = \boldsymbol{\mu}_1$. For GLMs, Simon (1973) showed that for nested hypotheses $\Omega_2 \subset \Omega_1 \subset \mathbb{R}^n$ with \mathbb{R}^n corresponding to the parameter space with an exact fit of $\tilde{\theta}_i$ and θ parameterized linearly in Ω_1 and Ω_2 ,

$$KL(\mathbf{y}, \boldsymbol{\mu}_2) = KL(\mathbf{y}, \boldsymbol{\mu}_1) + KL(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \quad (4)$$

when plugging in the values of maximum likelihood (ML) estimates for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. It is easy to see that (4) reduces to the ANOVA decomposition in linear models when $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ correspond to the linear or polynomial fit and the intercept-only model respectively, and the terms in (4) becomes SS total, SS residual, and SS regression respectively.

A linear or polynomial function of x may be restrictive and one may replace the linear dependence by a nonparametric function:

$$G\{\mu(x)\} = m(x). \quad (5)$$

Fan et al. (1995) discussed estimating $m(\cdot)$ by maximizing a locally weighted likelihood with a local polynomial approximation. Based on a Taylor's expansion at x , $\theta_i \approx \beta_0 + \beta_1(x_i - x) + \cdots + \beta_p(x_i - x)^p \stackrel{\text{def}}{=} \theta_i(x)$. This local polynomial approximation is plugged in the locally weighted log-likelihood $\ell_x(\mathbf{y}, \boldsymbol{\theta}_x) \stackrel{\text{def}}{=} \sum_i \ell\{y_i, \theta_i(x)\} K_h(x_i - x)$, where $\boldsymbol{\theta}_x = (\theta_1(x), \dots, \theta_n(x))^\top$, and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^\top$ maximizing $\ell_x(\mathbf{y}, \boldsymbol{\theta}_x)$ is solved. From the Taylor's expansion, $j! \hat{\beta}_j$ estimates $m^{(j)}(x)$, $j = 0, \dots, p$, which is $\theta^{(j)}(x)$ with the canonical link. Fan et al. (1995) derived asymptotic properties of $\hat{\beta}_j(x)$'s and adopted $G^{-1}\{\hat{\beta}_0(x)\}$ as an estimate for $\mu(x)$.

We will derive a local ANODev by adapting (4) for locally weighted likelihood. Define $\hat{\theta}_i(x) = \hat{\beta}_0 + \cdots + \hat{\beta}_p(x_i - x)^p$, an estimate of θ_i at x , $\hat{\boldsymbol{\theta}}_x = (\hat{\theta}_1(x), \dots, \hat{\theta}_n(x))^\top$, $\hat{\boldsymbol{\mu}}_x(x_i) = G^{-1}\{\hat{\theta}_i(x)\}$, and $\hat{\boldsymbol{\mu}}_x = (\hat{\mu}_x(x_1), \dots, \hat{\mu}_x(x_n))^\top$. As the $\hat{\beta}_j$'s maximize $\ell_x(\mathbf{y}, \boldsymbol{\theta}_x)$, the following equations hold:

$$\begin{aligned} \sum_i y_i(x_i - x)^j K_h(x_i - x) &= \sum_i \hat{\mu}_x(x_i)(x_i - x)^j K_h(x_i - x), & j = 0, \dots, p, \\ \sum_i y_i \hat{\theta}_i(x) K_h(x_i - x) &= \sum_i \hat{\mu}_x(x_i) \hat{\theta}_i(x) K_h(x_i - x). \end{aligned} \quad (6)$$

The last equation indicates that $(\mathbf{y} - \hat{\boldsymbol{\mu}}_x)$ is orthogonal to $\hat{\boldsymbol{\theta}}_x$ in the locally weighted inner product space with weights $K_h(x_i - x)$. Hence the fact of residuals being orthogonal to fitted values in ordinary linear models now becomes the fact of local residuals $(\mathbf{y} - \hat{\boldsymbol{\mu}}_x)$ being orthogonal to locally fitted canonical parameters $\hat{\boldsymbol{\theta}}_x$. For $\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x)$, an expression mimicking (2) for local deviance at x is therefore:

$$d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) = 2\{\ell_x(\mathbf{y}, \tilde{\boldsymbol{\theta}}) - \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x)\} = 2 \sum_i [y_i\{\tilde{\theta}_i - \hat{\theta}_i(x)\} - b(\tilde{\theta}_i) + b(\hat{\theta}_i(x))] K_h(x_i - x). \quad (7)$$

Though (7) is a natural definition from local likelihood, we are not aware of a similar quantity to (7) in the literature. Published works focused on global deviance by taking (2) with $G^{-1}\{\hat{\beta}_0(x_i)\}$ as estimates. In comparison, the local deviance (7) makes use of all coefficients $\hat{\beta}_0, \dots, \hat{\beta}_p$ and may be more sensible. Then (4) is adapted to form a local ANODev, and a global ANODev may be obtained by integrating local counterparts, as given in the following Theorem.

Theorem 1 (a) *A local ANODev is*

$$d_x(\mathbf{y}, \bar{y}) = d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) + d_x(\hat{\boldsymbol{\mu}}_x, \bar{y}), \quad (8)$$

where \bar{y} is the sample mean of \mathbf{y} , $d_x(\mathbf{y}, \bar{y})$ is (7) with $\hat{\boldsymbol{\mu}}_x$ and $\hat{\boldsymbol{\theta}}_x$ replaced by \bar{y} and $G(\bar{y})$ respectively, and

$$d_x(\hat{\boldsymbol{\mu}}_x, \bar{y}) \equiv 2\mathbf{E}_{\hat{\boldsymbol{\mu}}_x} \left[\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) - \ell_x\{\mathbf{y}, G^{-1}(\bar{y})\} \right] = 2 \left[\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) - \ell_x\{\mathbf{y}, G^{-1}(\bar{y})\} \right], \quad (9)$$

with the last expression holds based on (6).

(b) *A global ANODev is obtained by integrating the counterparts in (8):*

$$\int d_x(\mathbf{y}, \bar{y}) dx = \int d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) dx + \int d_x(\hat{\boldsymbol{\mu}}_x, \bar{y}) dx, \quad (10)$$

and $\int d_x(\mathbf{y}, \bar{y}) dx = KL(\mathbf{y}, \bar{y}) = D(\mathbf{y}, \bar{y})$ under a boundary condition that the weights $\int K_h(x_i - x) dx = 1$ for all x_i 's.

It is easy to show (8) due to (9) and hence the proof is omitted. Alternatively the proof in Simon (1973) for (4) can be adapted with kernel weights to show (8). Note that the local ANODev (8) holds in finite-sample cases for local polynomial estimates, similar to (4). For the Normal distribution with an identity link, (8) and (10) become the local and global ANOVA decompositions respectively in Huang and Chen (2008). They demonstrate that it is necessary to utilize all coefficients $\hat{\beta}_j, j = 0, \dots, p$, for the ANOVA decompositions to hold. Similarly, (8) and (10) hold with $\hat{\theta}_i(x)$ and $\hat{\mu}_x(x_i)$ based on the whole polynomial estimate, while most work in the literature focused only on $\hat{\beta}_0$.

As a by-product, the above derivations give rise to new estimators for θ_i 's and μ_i 's:

$$\theta_i^* = \int \hat{\theta}_i(x) K_h(x_i - x) dx \quad \text{and} \quad \mu_i^* = G^{-1}(\theta_i^*). \quad (11)$$

They are different from local estimates at x_i : $\hat{\beta}_0(x_i)$ and $G^{-1}\{\hat{\beta}_0(x_i)\}$. The asymptotic properties of θ_i^* and μ_i^* for interior points x_i with $p = 1$ and 3 are discussed in the following Remark.

Remark 1: Suppose that conditions (A1)-(A4) in the Appendix hold. Assume that $h \rightarrow 0$ and $nh^3 \rightarrow \infty$ as $n \rightarrow \infty$. We first define the ‘‘interior’’ and ‘‘boundary’’ points. For a kernel function defined on $[-1, 1]$, if the support of x_i 's is $[a, b]$, then define the interior region as $[a + 2h, b - 2h]$. See Huang and Chan (2013) for discussion on why the interior is not the conventional $[a + h, b - h]$. Then for interior points x_i with $p = 1$ and 3 , (a) the order of the asymptotic bias of θ_i^* is smaller than the conventional order $h^{(p+1)}$; i.e., the $h^{(p+1)}$ term of the bias of θ_i^* is zero; (b) the asymptotic variance of θ_i^* is of order $n^{-1}h^{-1}$; (c) the asymptotic behavior of μ_i^* is analogous to θ_i^* : the order of the bias is smaller than the conventional order $h^{(p+1)}$ and the asymptotic variance of μ_i^* is of order $n^{-1}h^{-1}$. See the Appendix for an outline of proofs. In the Gaussian case with an identity link, Huang and Chan (2013) show that the bias of θ_i^* for interior points is of order $h^{2(p+1)}$ for $p = 0, 1, 2, 3$, which is consistent with intuition that the higher the p , the smaller the order of the bias. Since the second-order expansions of the bias of $\hat{\beta}_j(x)$ for GLMs have not been addressed in the literature, we choose not to derive explicit bias expressions of θ_i^* in this paper. We thus focus on ANODev, while the issue of estimation may be studied in a future paper.

The following theorem concerns the asymptotic properties of integrated local likelihood $\int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) dx$.

Theorem 2 *Suppose that conditions (A1)-(A4) in the Appendix hold. Assume that $h \rightarrow 0$ and $nh^3 \rightarrow \infty$ as $n \rightarrow \infty$. For $p = 1$ and 3 ,*

(a) *the integrated likelihood function is asymptotically*

$$\int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) dx = \ell(\mathbf{y}, \boldsymbol{\theta}^*) + \mathcal{O}(h^{(p+1)}), \quad (12)$$

where $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_n^*)^\top$ with elements defined in (11);

(b) *the integrated deviance quantities are asymptotically*

$$\int d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) dx = KL(\mathbf{y}, \boldsymbol{\mu}^*) + \mathcal{O}(h^{(p+1)}) \quad \text{and} \quad \int d_x(\hat{\boldsymbol{\mu}}_x, \bar{y}) dx = KL(\boldsymbol{\mu}^*, \bar{y}) + \mathcal{O}(h^{(p+1)}), \quad (13)$$

where $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^\top$ with elements defined in (11).

The proof for Theorem 2 is given in the Appendix. The term $\int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) dx$ in (12) is a weighted integral of local likelihood with fitted local polynomials. In the literature, the idea of integrated likelihood was mentioned in Lehmann (2006), and Severini (2007) discussed integrated likelihood functions to eliminate nuisance parameters in parametric settings. To our knowledge, (12) has never been raised in the nonparametric regression literature, since the convention was to plug in $\hat{\beta}_0(x_i)$ in (1) for θ_i as an estimated likelihood function which ignores $\hat{\beta}_1, \dots, \hat{\beta}_p$. Theorem 2(b) together with Theorem 1(b) implies that

$$KL(\mathbf{y}, \bar{y}) = KL(\mathbf{y}, \boldsymbol{\mu}^*) + KL(\boldsymbol{\mu}^*, \bar{y}) + \mathcal{O}(h^{(p+1)}),$$

which shows that the classical ANODEV holds asymptotically in the local polynomial settings with $\boldsymbol{\mu}^*$, and hence (10) is more suitable for a global ANODEV as it holds in finite-sample cases. Based on integrated local likelihood, we next develop a chi-square test for examining significance of a nonparametric fit.

Theorem 3 *Under the conditions of Theorem 2, for testing $H_0 : m(x) = a_0$ with a_0 a constant, versus $H_a : m(x)$ is nonlinear and estimated by p -th order local polynomials with $p \geq 0$, the test statistic*

$$2 \left\{ \int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) dx - \ell(\mathbf{y}, \hat{a}_0) \right\} \quad (14)$$

is asymptotically distributed according to a χ^2 -distribution with a degree of freedom $\text{tr}(H_p^) - 1$, where \hat{a}_0 is the ML estimate under H_0 and H_p^* is the smoothing matrix for local p -th order polynomial regression defined in Huang and Chen (2008).*

More explicitly, H_p^* depending only on x_i 's, bandwidth h , and the kernel function $K(\cdot)$, is

$$H_p^* = \int W X_p (X_p^\top W X_p)^{-1} X_p^\top W dx, \quad (15)$$

where W is an n -dimensional diagonal matrix with $K_h(x_i - x)$ as its diagonal elements, and X_p is the $n \times (p + 1)$ design matrix with the $(j + 1)$ -th column $((x_1 - x)^j, \dots, (x_n - x)^j)^\top$, $j = 0, \dots, p$. The dependence of W and X_p on x is suppressed and the integration in (15) is performed element by element in the resulting matrix product. Huang and Chan (2013) discuss various properties of H_p^* , including being symmetric and shrinking.

3 Analysis of Deviance for Partial Linear Models

We extend the nonparametric ANODev in Section 2 to GPLMs by allowing

$$G(\mu) = \mathbf{z}^\top \alpha + m(x), \quad (16)$$

where \mathbf{z} is a K -dimensional covariate vector. Without loss of generality, the intercept is embedded in $m(\cdot)$. When y is Gaussian, PLMs are a popular class of semiparametric models, combining advantages of both linear and nonparametric regression. Huang and Davidson (2010) discussed ANOVA for PLMs and provided explicit expressions of the ANOVA SS quantities based on local polynomial regression. They showed that the SS residuals under PLMs can be decomposed into non- and parametric parts and subsequently developed F -tests using the differences in SS, similar to classical F -tests.

For GPLMs (16), if α is known or given, then the ANODev inference in Section 2 can be extended straightforwardly. A local and a global ANODev can be constructed as (8) and (10) with $\hat{\mu}_x(x_i) = G^{-1}\{\hat{\theta}_i(x)\}$, where $\hat{\theta}_i(x) = \mathbf{z}_i^\top \alpha + \hat{\beta}_0 + \cdots + \hat{\beta}_p(x_i - x)^p$ with local estimates $\hat{\beta}_j$'s under (16). To avoid confusion with the notation in Section 2, from now on $\hat{\mu}_x(x_i)$, $\hat{\theta}_i(x)$, $\hat{\boldsymbol{\theta}}_x$, $\hat{\boldsymbol{\mu}}_x$, $\boldsymbol{\mu}^{**}$, and $\boldsymbol{\theta}^{**}$ denote the estimates under GPLMs (16). When α is unknown, estimation of α can be done via a two-step ML procedure that updates the linear and nonparametric estimates iteratively, as discussed in Carroll et al. (1997). Since $\hat{\beta}_j$'s maximize the local likelihood, the equations in (6) continue to hold with $\hat{\boldsymbol{\theta}}_x$ and $\hat{\boldsymbol{\mu}}_x$ under (16). The interpretation that $(\mathbf{y} - \hat{\boldsymbol{\mu}}_x)$ is orthogonal to $\hat{\boldsymbol{\theta}}_x$ in the locally weighted inner product space with weights $K_h(x_i - x)$ continues to hold under GPLMs. For estimating α , ML yields

$$\sum_i y_i z_{ik} = \sum_i z_{ik} \int \hat{\mu}_x(x_i) K_h(x_i - x) dx, \quad k = 1, \dots, K, \quad (17)$$

where z_{ik} denotes the value of the k -th covariate for the i -th observation.

Theorems 1 and 2 are extended to GPLMs in the following as Theorem 4(a) and 4(b) respectively with $\hat{\boldsymbol{\mu}}_x$ and $\boldsymbol{\mu}^{**}$ when α is estimated by ML.

Theorem 4 (a) *The local and global ANODev (8) and (10) respectively hold for GPLMs with $\hat{\mu}_x(x_i)$ and $\hat{\boldsymbol{\mu}}_x$ when α is estimated by ML.*

(b) *Under the conditions in Theorem 2 and Condition (A5) in the Appendix, assume that α is estimated with a root- n rate by ML. For $p = 1$ or 3 , the expression in (12) holds*

with $\hat{\boldsymbol{\theta}}_x$ and $\boldsymbol{\theta}^{**}$ except the $\mathcal{O}(h^{p+1})$ term replaced by $\mathcal{O}(h^{p+1} + n^{-1/2})$. Similarly, (13) holds for GPLM with $\hat{\boldsymbol{\mu}}_x$ and $\boldsymbol{\mu}^{**}$ and the $\mathcal{O}(h^{p+1})$ terms replaced by $\mathcal{O}(h^{p+1} + n^{-1/2})$. Moreover $\int \hat{\boldsymbol{\mu}}_x(x_i)K_h(x_i - x)dx = \boldsymbol{\mu}_i^{**} + \mathcal{O}(h^{p+1} + n^{-1/2})$ and $(\mathbf{y} - \boldsymbol{\mu}^{**})$ is asymptotically orthogonal to the column space spanned by \mathbf{z} from (17).

(c) When the same kernel function and bandwidth are used in (5) and (16), the nonparametric model (5) is nested in (16). Then the local residual deviance from fitting (16) when $p \geq 0$ can be expressed as

$$d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) = d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) - d_x(\hat{\boldsymbol{\mu}}_x, \hat{\boldsymbol{\mu}}_x). \quad (18)$$

The proof of Theorem 4(b) is analogous to Theorem 2 and is thus omitted. We briefly outline the proof for Theorem 4(c). Based on (6) under GPLMs, we have $\sum_i y_i(x_i - x)^j K_h(x_i - x) = \sum_i \hat{\boldsymbol{\mu}}_x(x_i)(x_i - x)^j K_h(x_i - x), j = 0, \dots, p$. Then multiplying the j -th equation by $\hat{\beta}_j$ and summing them up, $\sum_i \{y_i - \hat{\boldsymbol{\mu}}_x(X_i)\} \hat{\boldsymbol{\theta}}_i(x) K_h(X_i - x) = 0$ is obtained and (18) is proved.

Theorem 4(c) implies that the local residual deviance for a GPLM is the deviance for fitting the nonparametric (5) minus a term due to the parametric component. That is, the reduction of local deviance after adding the parametric component can be expressed explicitly. A similar interpretation holds at a global scale after integrating the local counterparts of (18):

$$\int d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) dx = \int d_x(\mathbf{y}, \hat{\boldsymbol{\mu}}_x) dx - \int d_x(\hat{\boldsymbol{\mu}}_x, \hat{\boldsymbol{\mu}}_x) dx. \quad (19)$$

Analogous to Theorem 3, testing whether $m(\cdot)$ is significantly different from a constant may be based on the ratio of integrated local likelihoods.

Theorem 5 *Under model (16), assume that Conditions (A) in the Appendix hold and the data matrix Z for covariates \mathbf{z} is orthogonal to $\mathbf{x} = (x_1, \dots, x_n)^\top$ and the intercept column. For testing $H_0 : m(x) = a_0$ with a_0 a constant, versus $H_a : m(x)$ is nonlinear and estimated by p -th order local polynomials with $p \geq 0$, the test statistic*

$$2 \left\{ \int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) dx - \ell(\mathbf{y}, \hat{\boldsymbol{\alpha}}_0) \right\} \quad (20)$$

is asymptotically distributed according to a χ^2 -distribution with degrees of freedom $tr(H_p^) - 1$, where $\hat{\boldsymbol{\alpha}}_0 = (\hat{a}_0, \hat{\boldsymbol{\alpha}}^\top)^\top$ with \hat{a}_0 and $\hat{\boldsymbol{\alpha}}$ the ML estimates under H_0 .*

The assumption that Z is orthogonal to \mathbf{x} in Theorem 5 is required, in the sense that the corresponding off-diagonal elements of the local Fisher information is 0, for convenience

of deriving the asymptotic χ^2 -distribution of the test statistic (20). The proposed tests in Theorems 3 and 5 will depend on the bandwidth h , like the other nonparametric tests. We remark that the proposed tests are different from those in Hastie and Tibshirani (1990). The proposed test statistics utilize integrated likelihood that combines all maximized local likelihoods with fitted local polynomials, while the existing methods use only $G^{-1}(\hat{\beta}_0)$'s and the resulting expression is not based on maximized likelihood as $\hat{\beta}_1, \dots, \hat{\beta}_p$ are ignored.

Some work in the literature, e.g. Härdle et al. (1998), has considered testing whether the estimate of m in (16) is significantly different from a linear trend, $G(\mu) = \mathbf{z}^\top \alpha + a_0 + a_1 x$. However, the extension of Theorem 5 to testing a linear trend is non-trivial due to the fact that the variance function of y is allowed to be a function of the mean of y in GPLMs, which complicates the integrated likelihood term in (20). When the variance function of y is a constant, such as in classical linear models with the Gaussian assumption, an ANOVA-based F -type test is derived in Huang and Su (2009).

4 Simulation Results

We examine the empirical type-I error and power for the integrated likelihood ratio tests in Theorems 3 and 5. The simulation study focuses on logistic regression as we want to evaluate the proposed methods in order to analyze the German Bundesbank data. The sample size is $n = 200$ and 500 simulated data sets are generated. Local linear smoothing with the Epanechnikov kernel is used throughout this section. Critical values are taken from the χ^2 -distribution with 5% significance level.

Example 1:

$$\text{logit}(p) = -1.0 + ax^2,$$

where $a = 0, -1, -1.5, -2$, and $x \sim U(0, 1)$. The values of bandwidth $h = 0.1, 0.15, 0.2$, and 0.25 are used, which results in an average degrees of freedom 11.08, 7.69, 6.00, and 4.99 respectively. The proportion of rejection is given in Table 1. The type-I errors when $a = 0$ are below 0.05, indicating a reasonable approximation of the χ^2 distribution. When $a > 0$, we observe that the proportion of rejections increases as the value of the bandwidth increases

Example 2:

$$\text{logit}(p) = -1.0 + a \cos(2\pi x),$$

where $a = 0.5, 0.75, 1.0$, $x \sim U(0, 1)$. The degrees of freedom in this example are similar to Example 1, as the same values of the bandwidth are used and the distribution of the design points is identical to Example 1. This alternative family contains weak to strong nonlinear components and the proportion of rejections is given in Table 2. The behavior that the proportion of rejections increases as the value of the bandwidth increases under alternatives is observed again here.

Next we examine the performance of (20) under GPLMs.

Example 3:

$$\text{logit}(p) = b_1 z_1 + b_2 z_2 + a \exp(-16x^2),$$

where z_1 is first generated as binary taking values -1 and 1 with equal probabilities, $z_2 \sim N(0, 0.5^2)$, and $x \sim U(-0.5, 1)$. To satisfy the conditions in Theorem 5, z_1 and z_2 are then made orthogonal to \mathbf{x} and the intercept vector by multiplying $(I - P_1)$ where P_1 is the projection matrix for \mathbf{x} and intercept. After the orthogonalized z_1 and z_2 are obtained, $b_1 = 1$, $b_2 = -1$, and $a = 0, 1, 2, 3$. The average degrees of freedom corresponding to the values of $h = 0.15, 0.2$, and 0.25 are 11.17, 8.61, and 7.07 respectively. The proportion of rejections is given in Table 3. We observe that when $a = 0$ the type-I errors are below 5%, same as in Example 1. Table 3 shows satisfactory behavior of the proposed test for GPLMs.

Example 4:

$$\text{logit}(p) = b_1 z_1 + b_2 z_2 + a \cos(2\pi x),$$

where the data generation of z_1 , z_2 , and x is identical to Example 3, b_1 and b_2 are the same as Example 3, and $a = 1, 1.5, 2$. The results in Table 4 confirms with previous examples that the integrated likelihood ratio test perform reasonably well and the asymptotic χ^2 -distribution approximates well in finite-sample cases.

5 German Bundesbank Data

We apply the methodology in Section 3 to the German Bundesbank Data in year 2002, which was provided by CRC 649, Humboldt University, Berlin. The data contains 6123 firms in which 186 were insolvent. Each firm is described by a set of financial statement variables and those of insolvent firms are collected two years prior to insolvency. In accordance with literature, 28 ratios, x_1, \dots, x_{28} have been selected for the bankruptcy analysis. To ensure

the value of some variables as the denominator when calculating the ratios should not be zero, 2079 firms were retained with 92 insolvent. See Chen et al. (2011) for detail descriptions about the data. Based on support vector machines, Chen et al. (2011) selected x_{24} (account payable turnover) as the best predictor, and subsequently selected x_3 (OI/TA), x_{15} (CASH/TA), x_{12} (TL/TA), x_{26} (percentage of incremental inventories), x_{22} (inventory turnover), x_5 (EBIT/TA), and x_2 (net profit margin). We found x_3 and x_5 have a large sample correlation coefficient 0.95 and thus x_5 is removed in our analysis and we further include x_{25} (firm size) as it is shown to be an important variable on the probability of bankruptcy in the literature. Among the 8 predictors, there are two activity ratios (x_{24} and x_{22}), 2 profitability ratios (x_3 and x_2), one liquidity ratio (x_{15}), one leverage ratio (x_{12}), one percentage of change ratio (x_{26}), and one size variable (x_{25}).

Since x_{24} was selected as the most important predictor by Chen et al. (2011), we model its effects nonparametrically, while retaining linear trends for the remaining predictors in a logit model. Further examinations of x_{24} indicates that most values lie in $[0, 0.5]$ with only 15 observations in $(0.5, 20.52)$. In the region where data are sparse, smoothing methods will not be applicable. Thus we remove those 15 observations and the final sample size is 2064, in which 91 are insolvent. Local linear smoothing with the Epanechnikov kernel is used with $h = 0.125, 0.1$, and 0.075 corresponding to degrees of freedom 4.94, 5.92, and 7.59 respectively. The curves shown in Figure 1 clearly demonstrate a nonlinear trend of x_{24} . When $x_{24} < 0.3$ approximately, the default probability increases as x_{24} increases. Then the probability decreases for $0.03 < x_{24} < 0.04$, and increases again for $x_{24} > 0.4$. The proposed integrated likelihood ratio test gives p -values of 0.0045, 0.0072, and 0.015 respectively, indicating significance of this variable. Thus an interpretation in Chen et al. (2011) that a German firm is likely to go bankrupt when it has high turn over is not entirely correct; in our analysis, the effects of x_{24} on bankrupt probability is nonlinear.

In technical terms, accounts payable turnover ratio is an accounting liquidity metric that evaluates how fast a company pays off its creditors (suppliers). A high ratio means there is a relatively short time between purchase of goods and services and payment for them. Conversely, a lower accounts payable turnover ratio usually signifies that a company is slow in paying its suppliers. Since the y -axis in Figure 1 is in the negative range, it indicates that including x_{24} in the model has a protective effect. Taking the curve with $h = 0.1$ in Figure

1, when x_{24} increases from 0.1 to 0.3, the nonlinear estimate increases about 1.895, which means the odds ratio for a firm with $x_{24} = 0.3$ to become insolvent is $\exp(1.895) = 6.653$ times relative to that for a firm with $x_{24} = 0.1$. On the other hand, between $x_{24} = 0.3$ and 0.4, the nonlinear estimate decreases by an amount of -0.607 , implying that the odds ratio for a firm with $x_{24} = 0.4$ to become insolvent is $\exp(-0.607) = 0.545$ times relative to that for a firm with $x_{24} = 0.3$. Thus our analysis gives new insight that a German firm is likely to go bankrupt when it has higher turnover for roughly 97.5% of firms (0.3 is approximately 97.5-percentile of x_{24}), but for those firms with $0.3 < x_{24} < 0.4$ (approximately 97.5- to 99-percentile), the default probability decreases as x_{24} increases.

Appendix

The following conditions are similar to those in Fan *et al.* (1995).

CONDITIONS (A).

- (A1). The kernel $K(\cdot)$ is a Lipschitz continuous, bounded and symmetric probability density function, having a support on a compact interval, say $[-1, 1]$.
- (A2). The function $(\partial^2/\partial x^2)\ell\{y, \theta(x)\} < 0$ for $x \in R$ and y in the range of the response variable.
- (A3). The functions L' , $\theta^{(p+1)}$, $b''(\theta(\cdot)) \equiv V(\cdot)$, V'' , and $G^{(3)}$ are continuous.
- (A4). For each x in the support of $\{x_i, i = 1, \dots, n\}$, $V(x)$ and $G'(\mu(x))$ are nonzero.
- (A5). For model (16), the covariate vector \mathbf{z} is assumed to have a bounded support.

Proof of Remark 1: When $p = 1$, the bias of θ_i^* is expressed as follows:

$$\begin{aligned} \mathbb{E} \left\{ \int (\hat{\beta}_0(x) + (x_i - x)\hat{\beta}_1(x))K_h(x_i - x)dx \right\} - \theta_i = \\ \mathbb{E} \left[\int \{(\hat{\beta}_0(x) - \beta_0(x)) + (x_i - x)(\hat{\beta}_1(x) - \beta_1(x))\}K_h(x_i - x)dx - \right. \\ \left. \int \{\beta_2(x)(x_i - x)^2 + r(x, x_i)\}K_h(x_i - x)dx \right], \end{aligned} \quad (21)$$

where $r(x, x_i)$ denotes the remainder terms. Plugging the first-order term of the asymptotic bias of $\hat{\beta}_0(x)$ (Fan *et al.* (1995) and Aerts and Claeskens (1997)) in (21), leads to cancellation

with the $\beta_2(x)$ -term in (21). The remaining term $\int (x_i - x)(\hat{\beta}_1(x) - \beta_1(x))K_h(x_i - x)dx$ is of order h^4 . Thus the h^2 -order term in (21) is zero. Similar arguments can be shown for $p = 3$.

Fan et al. (1995) and Aerts and Claeskens (1997) showed that the variance of $\hat{\beta}_j(x)$ is of order $n^{-1}h^{-2j-1}$ when p is odd. Then the variance of $\{\hat{\beta}_0(x) + (x_i - x)\hat{\beta}_1(x)\}$ is of order $n^{-1}h^{-1}$ and hence the variance of θ_i^* is of order $n^{-1}h^{-1}$. Finally, it is straightforward to show (c) based on (a) and (b) since $\mu_i^* = G^{-1}(\theta_i^*)$.

Proof of Theorem 2: We only need to show Theorem 2(a) while Theorem 2(b) follows directly from Theorem 2(a). For the left-hand side of (12), ignoring the $c(y, \phi)$ and $a(\phi)$ terms in (1) which is unrelated to x , the integrated likelihood is

$$\int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x)dx = \sum_i \left\{ y_i \int \hat{\theta}_i(x)K_h(x_i - x)dx - \int b(\hat{\theta}_i(x))K_h(x_i - x)dx \right\}.$$

By a Taylor's expansion,

$$b(\hat{\theta}_i(x)) = b(\theta_i^*) + b'(\theta_i^*)\{\hat{\theta}_i(x) - \theta_i^*\} + b''(\theta_i^*)\{\hat{\theta}_i(x) - \theta_i^*\}^2/2 + r_i(x),$$

where $r_i(x)$ denotes the remainder terms. Note that for the linear term,

$$\int b'(\theta_i^*)(\hat{\theta}_i(x) - \theta_i^*)K_h(x_i - x)dx = b'(\theta_i^*)(\theta_i^* - \theta_i^*) = 0.$$

The quadratic term $\int (\hat{\theta}_i(x) - \theta_i^*)^2 K_h(x_i - x)dx = \int \hat{\theta}_i(x)\{\hat{\theta}_i(x) - \theta_i^*\}K_h(x_i - x)dx$. For $\hat{\theta}_i(x) - \theta_i^* = (\hat{\theta}_i(x) - \theta_i) - (\theta_i^* - \theta_i)$, the first term $\hat{\theta}_i(x) - \theta_i = (\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)(x_i - x) + \dots + (\hat{\beta}_p - \beta_p)(x_i - x)^p + r'_i(x)$ by a Taylor's expansion of θ_i , where $r'_i(x)$ denotes the remainder terms. Then based on Theorem 1(a) of Fan et al. (1995), $\hat{\theta}_i(x) - \theta_i$ is of order $\mathcal{O}(h^{p+1})$. For the second term $\theta_i^* - \theta_i$, it is of order $\mathcal{O}(h^{p+1})$ based on Remark 1. Thus $\int b(\hat{\theta}_i(x))K_h(x_i - x)dx = b(\theta_i^*) + \mathcal{O}(h^{p+1})$ and (12) is proved.

Proof of Theorem 3: Let $\ell(y_i, a_0)$ be the likelihood corresponding to y_i with $\theta_i = a_0$. Define $\ell_x(\mathbf{y}, a_0) = \sum_i \ell(y_i, a_0)K_h(x_i - x)$, and it is clear that $\int \ell_x(\mathbf{y}, a_0)dx = \ell(\mathbf{y}, a_0)$. Recall $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^\top$ maximizes local likelihood at x with local polynomial approximation. Expanding $\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x)$, which is a function of $\hat{\boldsymbol{\beta}}$, around a $(p+1)$ -length vector $\mathbf{a}_0 = (a_0, 0, \dots, 0)^\top$,

$$\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) - \ell_x(\mathbf{y}, \mathbf{a}_0) = \left\{ \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top (\hat{\boldsymbol{\beta}} - \mathbf{a}_0) + \frac{1}{2} (\hat{\boldsymbol{\beta}} - \mathbf{a}_0)^\top \frac{\partial^2 \ell_x}{\partial \boldsymbol{\beta}^2}(\mathbf{y}, \mathbf{a}_0) (\hat{\boldsymbol{\beta}} - \mathbf{a}_0) + \mathcal{O}_p(\sqrt{n^{-1}h^{-1}}). \quad (22)$$

Substituting the expansion

$$\hat{\boldsymbol{\beta}} - \mathbf{a}_0 = i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) + \mathcal{O}_p(\sqrt{n^{-1}h^{-1}}),$$

where $i_x(\mathbf{a}_0) = \mathbb{E} \left\{ -\frac{\partial^2 \ell_x}{\partial \boldsymbol{\beta}^2}(\mathbf{y}, \boldsymbol{\theta}) \right\}(\mathbf{a}_0)$, we have for (22),

$$\begin{aligned} \left\{ \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) + \frac{1}{2} \left\{ \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top i_x(\mathbf{a}_0)^{-1} \frac{\partial^2 \ell_x}{\partial \boldsymbol{\beta}^2}(\mathbf{y}, \mathbf{a}_0) i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) \\ + \mathcal{O}_p(\sqrt{n^{-1}h^{-1}}). \end{aligned}$$

Since $\frac{\partial^2 \ell_x}{\partial \boldsymbol{\beta}^2}(\mathbf{y}, \mathbf{a}_0) = -i_x(\mathbf{a}_0) + \mathcal{O}_p(\sqrt{nh})$, it follows that $\int 2\{\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) - \ell_x(\mathbf{y}, a_0)\} dx$ is

$$\int \left\{ \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0) dx = \left\{ i(a_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, a_0) \right\}^\top H_p^* \left\{ i(a_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, a_0) \right\},$$

where the last expression is obtained by plugging in the explicit expressions of $\frac{\partial \ell_x}{\partial \boldsymbol{\beta}}(\mathbf{y}, \mathbf{a}_0)$ and $i_x(\mathbf{a}_0)^{-1}$, and $i(a_0)$ is the information matrix under H_0 .

By standard ML theory, $2\{\ell(\mathbf{y}, \hat{a}_0) - \ell(\mathbf{y}, a_0)\}$ is asymptotically distributed according to a χ^2 distribution with 1 degree of freedom. Hence the test statistic (14) becomes

$$\left\{ i(a_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, a_0) \right\}^\top \{H_p^* - P_0\} \left\{ i(a_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, a_0) \right\},$$

where P_0 is an $n \times n$ matrix with $1/n$ in all entries, i.e. the projection matrix from fitting a constant. From Huang and Chen (2008), H_p^* is a symmetric and asymptotically idempotent matrix. Note that $\left\{ i(a_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, a_0) \right\}$ is asymptotically normally distributed with mean vector 0 and identity covariance matrix. Since Huang and Chen (2008) have shown that $H_p^* - P_0$ is symmetric and asymptotically idempotent, the test statistic (14) has an asymptotic χ^2 -distribution with degrees of freedom $\text{tr}\{(H_p^*) - 1\}$.

Proof of Theorem 5:

The proof is an extension from that of Theorem 3. Let $\boldsymbol{\alpha}_0 = (a_0, \boldsymbol{\alpha}^\top)^\top$ denote the parameter vector under H_0 when $\theta_i = a_0 + \boldsymbol{\alpha}^\top \mathbf{z}_i$ and $\ell(\mathbf{y}, \boldsymbol{\alpha}_0)$ be the corresponding likelihood. Define local likelihood at x under H_0 by $\ell_x(\mathbf{y}, \boldsymbol{\alpha}_0) = \sum_i \ell(y_i, \boldsymbol{\alpha}_0) K_h(x_i - x)$ and hence $\int \ell_x(\mathbf{y}, \boldsymbol{\alpha}_0) dx = \ell(\mathbf{y}, \boldsymbol{\alpha}_0)$. We consider $\left\{ \int \ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) dx - \ell(\mathbf{y}, \boldsymbol{\alpha}_0) \right\}$ and $\{\ell(\mathbf{y}, \hat{\boldsymbol{\alpha}}_0) - \ell(\mathbf{y}, \boldsymbol{\alpha}_0)\}$ separately, whose difference becomes (14).

Let $\mathbf{b} = (\beta^\top, \alpha^\top)^\top$ be the parameter vector under H_1 . Expanding $\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x)$, which is a function of $\hat{\mathbf{b}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\alpha}}^\top)^\top$, around a $(p + K + 1)$ -length vector $\mathbf{a}_0 = (a_0, 0, \dots, 0, \alpha^\top)^\top$,

$$\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) - \ell_x(\mathbf{y}, \mathbf{a}_0) = \left\{ \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top (\hat{\mathbf{b}} - \mathbf{a}_0) + \frac{1}{2} (\hat{\mathbf{b}} - \mathbf{a}_0)^\top \frac{\partial^2 \ell_x}{\partial \mathbf{b}^2}(\mathbf{y}, \mathbf{a}_0) (\hat{\mathbf{b}} - \mathbf{a}_0) + \mathcal{O}_p(\sqrt{n^{-1}h^{-1}}). \quad (23)$$

Substituting the expansion under H_0

$$\hat{\boldsymbol{\beta}} - \mathbf{a}_0 = i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) + \mathcal{O}_p(\sqrt{n^{-1}h^{-1}}),$$

where $i_x(\mathbf{a}_0) = \mathbb{E} \left\{ -\frac{\partial^2 \ell_x}{\partial \mathbf{b}^2}(\mathbf{y}, \boldsymbol{\theta}) \right\}(\mathbf{a}_0)$, we have for (23),

$$\left\{ \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) + \frac{1}{2} \left\{ \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top i_x(\mathbf{a}_0)^{-1} \frac{\partial^2 \ell_x}{\partial \mathbf{b}^2}(\mathbf{y}, \mathbf{a}_0) i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) + \mathcal{O}_p(\sqrt{n^{-1}h^{-1}}).$$

Since $\frac{\partial^2 \ell_x}{\partial \mathbf{b}^2}(\mathbf{y}, \mathbf{a}_0) = -i_x(\mathbf{a}_0) + \mathcal{O}_p(\sqrt{nh})$ under H_0 , it follows that $\int 2\{\ell_x(\mathbf{y}, \hat{\boldsymbol{\theta}}_x) - \ell_x(\mathbf{y}, \boldsymbol{\alpha}_0)\} dx$ is asymptotically

$$\begin{aligned} & \int \left\{ \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) \right\}^\top i_x(\mathbf{a}_0)^{-1} \frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0) dx \\ &= \left\{ i(\boldsymbol{\alpha}_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\alpha}_0) \right\}^\top \{H_p^* + P_z\} \left\{ i(\boldsymbol{\alpha}_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\alpha}_0) \right\}, \end{aligned}$$

where P_z is the projection matrix for Z and $i(\boldsymbol{\alpha}_0)$ is the information matrix under H_0 . The last expression is obtained by plugging in the explicit expressions of $\frac{\partial \ell_x}{\partial \mathbf{b}}(\mathbf{y}, \mathbf{a}_0)$ and $i_x(\mathbf{a}_0)^{-1}$ and using the assumption that \mathbf{x} and Z are orthogonal. From Huang and Davidson (2010), $\{H_p^* + P_z\}$ is a symmetric and asymptotically idempotent matrix.

For the other term, $2\{\ell(\mathbf{y}, \hat{\boldsymbol{\alpha}}_0) - \ell(\mathbf{y}, \boldsymbol{\alpha}_0)\}$ is asymptotically distributed according to a χ^2 -distribution with $(K + 1)$ degree of freedom under H_0 by standard ML theory. Hence the test statistic (14) becomes

$$\left\{ i(\boldsymbol{\alpha}_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\alpha}_0) \right\}^\top \{H_p^* - P_0\} \left\{ i(\boldsymbol{\alpha}_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\alpha}_0) \right\},$$

where P_0 is the same as in the proof of Theorem 3. Note that $\left\{ i(\boldsymbol{\alpha}_0)^{-1/2} \frac{\partial \ell}{\partial \boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\alpha}_0) \right\}$ is asymptotically normally distributed with mean vector 0 and identity covariance matrix. Hence the test statistic (14) has an asymptotic chi-square distribution with degrees of freedom $\text{tr}\{(H_p^*) - 1\}$.

References

- Aerts, M., and Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *J. Amer. Statist. Assoc.*, 92, 1536-1545.
- Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* 92, 477-489.
- Chen, S., Härdle, W., and Moro, R. (2011). Modeling default risk with support vector machines. *Quantitative Finance*, 11, 135-154.
- Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall
- Fan, J., Heckman, N. E., and Wand, M. P. (1995). Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions. *J. Amer. Statist. Assoc.* 90, 141-150.
- Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.*, 29, 153-193.
- Green, P. J., and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*, London: Chapman and Hall.
- Härdle, W., Müller, M., and Mammen, E. (1998). Testing Parametric Versus Semiparametric Modeling in Generalized Linear Models. *J. Amer. Statist. Assoc.*, 93, 1461-1474.
- Härdle, W. K., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Berlin: Springer.
- Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models*, London: Chapman and Hall.
- Hastie, T. J., and Tibshirani, R. J. (1987). Local likelihood estimation. *J. Amer. Statist. Assoc.*, 82, 559-567.
- Huang, L.-S., and Chan, K.-S. (2013). Local polynomial and penalized trigonometric series regression. *Manuscript*, submitted.

- Huang, L.-S., and Chen, J. (2008). Analysis of Variance, Coefficient of Determination, and F-test for Local Polynomial Regression. *Ann. Statist.*, 36, 2085-2109.
- Huang, L.-S., and Davidson, P. W. (2010). Analysis of variance and F -tests for partial linear models with applications to environmental health data. *J. Amer. Statist. Assoc.*, 105, 991-1004.
- Huang, L.-S., and Su, H. (2009). Nonparametric F -tests for nested global and local polynomial models. *J. Statist. Plann. Inference*, 139, 1372-1380.
- Lehmann, E. L. (2006). On likelihood ratio tests. *Lecture Notes-Monograph Series Vol. 49, Optimality: The Second Erich L. Lehmann Symposium*, 1-8. Institute of Mathematical Statistics.
- Li, R., and Liang, H. (2008). Variable selection in semiparametric regression modeling. *Ann. Statist.*, 36, 261-286.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd Edition. London: Chapman and Hall.
- Severini, T. A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika*, 94, 529-542.
- Simon, G. (1973). Additivity of information in exponential family probability laws. *J. Amer. Statist. Assoc.*, 68, 478-482.

Table 1: Proportion of rejections for Example 1

Proportion of rejection	$h = 0.1$	$h = 0.15$	$h = 0.2$	$h = 0.25$
$a = 0$	0.048	0.040	0.036	0.030
$a = -1$	0.244	0.282	0.320	0.334
$a = -1.5$	0.558	0.602	0.640	0.674
$a = -2$	0.800	0.858	0.906	0.932

Table 2: Proportion of rejections for Example 2

Proportion of rejection	$h = 0.1$	$h = 0.15$	$h = 0.2$	$h = 0.25$
$a = 0.5$	0.286	0.348	0.370	0.384
$a = 0.75$	0.588	0.600	0.738	0.742
$a = 1.0$	0.892	0.926	0.946	0.946

Table 3: Proportion of rejections for Example 3

Proportion of rejection	$h = 0.15$	$h = 0.2$	$h = 0.25$
$a = 0$	0.034	0.026	0.022
$a = 1$	0.132	0.150	0.150
$a = 2$	0.620	0.654	0.682
$a = 3$	0.926	0.946	0.952

Table 4: Proportion of rejections for Example 4

Proportion of rejection	$h = 0.15$	$h = 0.2$	$h = 0.25$
$a = 1$	0.306	0.332	0.360
$a = 1.5$	0.685	0.732	0.752
$a = 2$	0.921	0.946	0.950

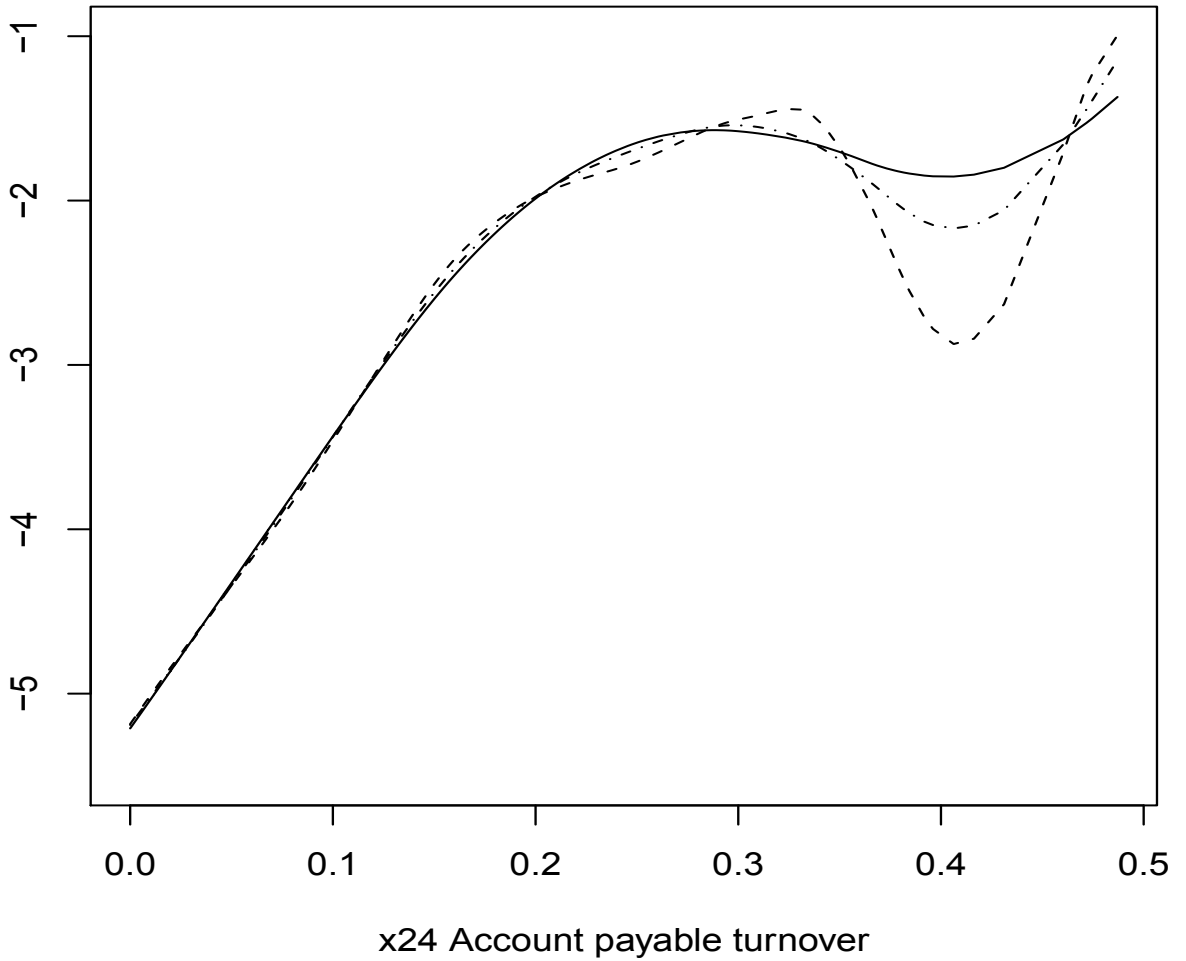


Figure 1: Plot of the nonlinear trends of x_{24} in predicting bankrupt probabilities with bandwidth $h = 0.125$ (solid line), $h = 0.1$ (dotdash line), and $h = 0.075$ (dashed line) for the 2002 German Bundesbank Data.

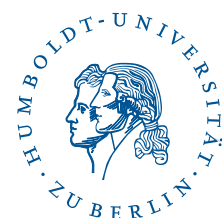
SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
- 002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
- 003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
- 004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
- 005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
- 006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.
- 007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.
- 008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.
- 009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.
- 010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.
- 011 "The Real Consequences of Financial Stress" by Stefan Mittnik and Willi Semmler, February 2013.
- 012 "Are There Bubbles in the Sterling-dollar Exchange Rate? New Evidence from Sequential ADF Tests" by Timo Bettendorf and Wenjuan Chen, February 2013.
- 013 "A Transfer Mechanism for a Monetary Union" by Philipp Engler and Simon Voigts, March 2013.
- 014 "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, March 2013.
- 015 "Cyclical Variation in Labor Hours and Productivity Using the ATUS" by Michael C. Burda, Daniel S. Hamermesh and Jay Stewart, March 2013.
- 016 "Quantitative forward guidance and the predictability of monetary policy – A wavelet based jump detection approach –" by Lars Winkelmann, April 2013.
- 017 "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency" by Markus Bibinger, Nikolaus Hautsch, Peter Malec and Markus Reiss, April 2013.
- 018 "Fair re-valuation of wine as an investment" by Fabian Y.R.P. Bocart and Christian M. Hafner, April 2013.
- 019 "The European Debt Crisis: How did we get into this mess? How can we get out of it?" by Michael C. Burda, April 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 020 "Disaster Risk in a New Keynesian Model" by Maren Brede, April 2013.
- 021 "Econometrics of co-jumps in high-frequency data with noise" by Markus Bibinger and Lars Winkelmann, May 2013.
- 022 "Decomposing Risk in Dynamic Stochastic General Equilibrium" by Hong Lan and Alexander Meyer-Gohde, May 2013.
- 023 "Reference Dependent Preferences and the EPK Puzzle" by Maria Grith, Wolfgang Karl Härdle and Volker Krätschmer, May 2013.
- 024 "Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations" by Hong Lan and Alexander Meyer-Gohde, May 2013.
- 025 "The 'Celtic Crisis': Guarantees, transparency, and systemic liquidity risk" by Philipp König, Kartik Anand and Frank Heinemann, May 2013.
- 026 "State Price Densities implied from weather derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Huei-Wen Teng, May 2013.
- 027 "Bank Lending Relationships and the Use of Performance-Sensitive Debt" by Tim R. Adam and Daniel Streitz, May 2013.
- 028 "Analysis of Deviance in Generalized Partial Linear Models" by Wolfgang Karl Härdle and Li-Shan Huang, May 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

