Göran Kauermann:

On a Small Sample Adjustment for the Profile Score Function in Semiparametric Smoothing Models

Sonderforschungsbereich 386, Paper 162 (1999)

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

On a Small Sample Adjustment for the Profile Score Function in Semiparametric Smoothing Models

Göran Kauermann *

Abstract

We consider the profile score function in models with smooth and parametric components. If local respectively weighted likelihood estimation is used for fitting the smooth component, the resulting profile likelihood estimate for the parametric component is asymptotically efficient as shown in Severini & Wong (1992). However, as in solely parametric models the profile score function is not unbiased. We propose a small sample bias adjustment which results by extending the correction suggested in McCullagh & Tibshirani (1990) to the framework of semiparametric models.

Keywords: Local Likelihood, Profile Likelihood, Semiparametric Models, Smoothing.

1 Introduction

We consider semiparametric models having two types of parameters, a finite dimensional component θ and a nonparametric smooth component $\varphi(\cdot)$. With (y, x) we denote a random vector where y is considered as response variable having x as vector of explanatory quantities. Given x we assume y to be distributed according to $y|x \sim f\{y|x; \theta, \varphi(u)\}$ where θ is a finite dimensional parameter and $\varphi(\cdot)$ is a smooth possibly multivariate function in u, with u being some metrically scaled covariate contained in x. A typical example is the semiparametric regression model where

*Institute for Statistics, Ludwigs-Maximilians University, Akademiestraße 1, D 80796 Munich, Germany; kauerman@stat.uni-muenchen.de

the expectation of the response variable y is modeled as $E(y|x) = \mu\{z\theta + \varphi(u)\}$ with $\mu(\cdot)$ as a known response function (or inverse link function) and z being some factorial covariate contained in x. The predictor $z\theta + \varphi(u)$ here consists of the unknown regression parameter θ, which serves as parameter of interest, accompanied by the unknown smooth function $\varphi(\cdot)$ taken as nonparametric nuisance component. Models of this type are discussed by several authors, see e.g. Heckman (1986), Speckman (1988), Severini & Wong (1992), Severini & Staniswalis (1994) or Hunsberger (1995).

Let (y_i, x_i), $i = 1, \ldots, n$, denote a random sample with $l_i(\theta, \varphi_i) = \log f(y_i|x_i; \theta, \varphi_i)$ as log likelihood contribution where $\varphi_i = \varphi(u_i)$. We denote with $\hat{\varphi}_i$ a smooth estimate of φ_i yielding

$$l^P(\theta) \quad = \quad \sum_i l_i(\theta, \hat{\varphi}_i) \tag{1}$$

as profile log likelihood function for θ. As in a solely parametric setting, this profile likelihood does not fulfill the usual likelihood properties. In particular the expectation of the resulting profile score function $l_\theta^P(\theta) = \partial l^P(\theta)/\partial\theta$ is not zero. We show that the resulting bias has order $O(h^{-p})$ where h is the smoothing parameter or bandwidth used for smoothing $\varphi(\cdot)$ and p is the dimension u. Hence for $h \to 0$ the bias of the profile score function is increasing. This is in contrast to solely parametric models where the bias of the profile score function typically has order $O(1)$. Our objective is to derive an approximation $b_\theta(\theta)$ for the dominating part of the bias such that the adjusted profile score function

$$l_\theta^{AP}(\theta) \quad = \quad l_\theta^P(\theta) - b_\theta(\theta) \tag{2}$$

is unbiased up to the second considered asymptotic order.

In a solely parametric setting various methods for adjusting profile likelihood functions have been suggested in the literature, see e.g. McCullagh (1987), Cox & Reid (1987), Davison (1988), McCullagh & Nelder (1989, ch. 7), Barndorff-Nielsen (1991) or Reid (1995). Corrections can be constructed by approximating the distribution of the estimated component in the profile likelihood function and then either marginalizing over or conditioning on the estimated parameters. McCullagh & Tibshirani (1990) suggest an additive adjustment based on an expansion of the profile score function. This concept is extended by DiCiccio, Martin, Stern & Young (1996) or Stern (1997) to generally adjust the bias of the second order derivative. Severini (1998) suggests an approximation of the profile likelihood function for multivariate but finite dimensional nuisance parameters. A major prerequisite of the adjustments above is that the nuisance parameter is estimated by the usual parametric $\sqrt{n}$ convergence. This rate of convergence can however not be achieved if the nuisance parameter is a smooth function. Here, the typical rate of convergence is $O_p(h^2) + O_p(n^{-1/2}h^{-p/2})$. The approach of McCullagh & Tibshirani (1990) however can be extended to the semiparametric framework, as demonstrated below.

2 Bias of the Profile Score Function

For estimation we use local likelihood for the smooth component $\varphi(\cdot)$ and profile likelihood estimation for the parametric component θ (see also Severini & Wong, 1992). Let w_{ij} denote some kernel weights $w_{ij} = K\{(u_i - u_j)/h\}/K(0)$, where $K(\cdot)$ is a multivariate positive, unimodal and symmetrical kernel function (see e.g. Staniswalis, 1989) and h as bandwidth. The weights are normed to have range $[0, 1]$ with maximal value 1 for $u_i \equiv u_j$. For fixed θ the estimate for $\varphi_i = \varphi(u_i)$ is now obtained by maximizing the local likelihood function

$$\sum_j w_{ij} l_j(\theta, \varphi_i) \tag{3}$$

with respect to φ_i. The resulting estimates $\hat{\varphi}_i$, $i = 1, \dots n$, thereby typically depends on θ, which is suppressed in the notation. Inserting the estimates $\hat{\varphi}_i$ in the likelihood function yields (1) as profile log likelihood function for θ.

The smoothing parameter h in (3) steers the amount of smoothing. It is assumed that h fulfills the standard conditions $h \to 0$ and $nh^p \to \infty$, where p is the dimension of u. We postulate $p \leq 3$ to ensure $\sqrt{n}$ convergence for the parametric component. Moreover for technical reasons we require that $nh^{4+p} \to 0$, which means that $\varphi(\cdot)$ is undersmoothed. This point is discussed in more detail later in the paper.

Let the components of θ be indexed by $a, b, ..$, i.e. $\theta = (\theta^a, \theta^b, \dots)$, while the letters $r, s, t, \dots$ are used to index the components of $\varphi_i = (\varphi_i^r, \varphi_i^s, \varphi_i^t, \dots)$, $i = 1, \dots, n$. With subscripts we denote derivatives, e.g. $l_{i;r}(\theta, \varphi_i) = \partial l_i(\theta, \varphi_i)/\partial \varphi_i^r$ or $l_{i;a}(\theta, \varphi_i) = \partial l_i(\theta, \varphi_i)/\partial \theta^a$. If the derivatives are evaluated at the true parameter value we drop the corresponding arguments, e.g. we write $l_{i;r}$ for $l_{i;r}(\theta, \varphi_i)$. Similarly we write $\hat{l}_{i;r}$ for $l_{i;r}(\theta, \hat{\varphi}_i)$ when θ is the true parameter. Finally we denote the cumulants of the likelihood by $\kappa_{i;r,s} = -E(l_{i;rs}) = E(l_{i;r} l_{i;s})$ or $\kappa_{i;rs,t} = E(l_{i;rs} l_{i;t})$, for instance.

Employing this notation allows to derive the profile score function as

$$\frac{\partial l^P(\theta)}{\partial \theta^a} = l_a^P(\theta) = \sum_i l_{i;a}(\theta, \hat{\varphi}_i) + \hat{\varphi}_{i;a}^r l_{i;r}(\theta, \hat{\varphi}_i) \tag{4}$$

where $\hat{\varphi}_{i;a}^r = \partial \hat{\varphi}_i^r / \partial \theta_a$ and Einstein's summation convention implies that we sum over repeated sub- and superscripts. The second component in (4) vanishes in a solely parametric framework or if the estimates $\hat{\varphi}_i$ do not depend on θ. In general however the profile score function in semiparametric models has two components.

For the calculation of the bias $b_a(\theta)$ we expand (4) to

$$l_a^P(\theta) \quad = \quad \sum_i l_{i;a} + \sum_i l_{i;ar}(\hat{\varphi}_i^r - \varphi_i^r) + \sum_i \hat{\varphi}_{i;a}^r \{l_{i;r} + l_{i;rs}(\hat{\varphi}_i^s - \varphi_i^s)\} + \dots . \tag{5}$$

Moreover, expansions for $(\hat{\varphi}_i^r - \varphi_i^r)$ and $\hat{\varphi}_{i;a}^r$ are required to calculate the bias of (5). An expansion for $(\hat{\varphi}_i^r - \varphi_i^r)$ is found from the local estimating equation

$$0 \quad = \quad \sum_j w_{ij} l_{j;r}(\theta, \hat{\varphi}_i) = \sum_j w_{ij} l_{j;r}(\theta, \varphi_i) + \sum_j w_{ij} l_{j;rs}(\theta, \varphi_i)(\hat{\varphi}_i^s - \varphi_i^s) + \dots . \tag{6}$$

Solving (6) for $(\hat{\varphi}_i^r - \varphi_i^r)$ provides an asymptotic expansion for $(\hat{\varphi}_i^r - \varphi_i^r)$. One should however note that the likelihood contributions in (6) are not unbiased, i.e. $E(l_{j;r}(\theta, \varphi_i)) \neq 0$, since φ_i and φ_j can differ. We therefore decompose the likelihood contributions to

$$\begin{aligned} l_{j;r}(\theta, \varphi_i) &= U_{j;r} + \delta_{j,i;r} \\ l_{j;rs}(\theta, \varphi_i) &= -\kappa_{i;rs} + U_{j;rs} + \delta_{j,i;rs} \end{aligned}$$

and so on. Components denoted by U are now standard likelihood terms with zero mean, i.e. $U_{j;r} = l_{j;r}$ or $U_{j;rs} = l_{j;rs} + \kappa_{i;r,s}$. Smoothing bias terms are collected in δ, e.g. $\delta_{j,i;r} = l_{j;rs}(\varphi_i^s - \varphi_j^s) + l_{j;rst}(\varphi_i^s - \varphi_j^s)(\varphi_i^t - \varphi_j^t)/2 + \dots$. Moreover we denote with $n_i = \sum_j w_{ij}$ the "local sample size" and we use the bar notation to define locally weighted means, e.g.

$$\bar{U}_{i;r} = n_i^{-1} \sum_j w_{ij} U_{j;r} \ , \quad \bar{\delta}_{i;r} = n_i^{-1} \sum_j w_{ij} \delta_{j,i;r} \ , \quad \bar{\kappa}_{i;r,s} = n_i^{-1} \sum_j w_{ij} \kappa_{j;r,s} .$$

Asymptotic consideration of the components above require some regularity conditions. First, we assume that the observed values of u_i become infinitely dense for growing sample size on some bounded support $\mathcal{U}$, say. This ensures, for instance, that $n_i = O(nh^p)$ and in turn $\bar{U}_{i;r} = O_p(n_i^{-1/2}) = O_p(n^{-1/2} h^{-p/2})$. Moreover, the information about $\varphi(\cdot)$ is supposed to grow sufficiently fast with increasing sample size such that $\bar{\kappa}_{i;r,s} = O(1)$ and $\bar{\kappa}_i^{r,s} = O(1)$, where $\bar{\kappa}_i^{r,s}$ is the matrix inverse of $\bar\kappa_{i;r,s}$. The same rate of convergence is also assumed for all higher order cumulants. Finally, the functional form of $\varphi(\cdot)$ is assumed to be sufficiently smooth, i.e. at least two times continuously differentiable, which controls the asymptotic order of the smoothing bias terms. In particular this yields that δ components are dominated by $O(h^2)$, e.g. $E(\bar\delta_{i;r}) = O(h^2)\{1+O_p(n_i^{-1/2})\}$. The asymptotic orders mentioned above hold only if u_i is an inner point of $\mathcal{U}$, and boundary points typically show weaker convergence. The fraction of boundary points however is of asymptotically negligible order and therefore we generally neglect boundary effects in what follows.

The provided notation can now be used to invert (6) which gives

$$\begin{aligned}\hat\varphi_i^r - \varphi_i^r &= \bar\kappa_i^{r,s}\,(\bar U_{i;s} + \bar\delta_{i;s}) + \bar\kappa_i^{r,s}\,\bar\kappa_i^{t,u}\,(\bar U_{i;st} + \bar\delta_{i;st})(\bar U_{i;u} + \bar\delta_{i;u}) \\ &\quad + \frac{1}{2}\bar\kappa_i^{r,s,t}\,(\bar U_{i;s} + \bar\delta_{i;s})(\bar U_{i;t} + \bar\delta_{i;t}) \\ &\quad + O(h^6) + O_p(n_i^{-1/2}h^4) + O_p(n_i^{-1}h^2) + O_p(n_i^{-3/2})\end{aligned} \tag{7}$$

where $\bar\kappa_i^{r,s,t} = \bar\kappa_i^{r,u}\,\bar\kappa_i^{s,v}\,\bar\kappa_i^{t,w}\,\bar\kappa_{i;uvw}$. The correction terms in (7) are found by simple calculations, see the appendix for further details. Note that formula (7) is conspicuously similar to expansions for standard likelihood functions, as for instance found in McCullagh (1987, p. 209).

For the calculation of the bias of (5) we also need the expansion of $\hat\varphi_{i;a}^r$. This can be obtained by differentiating (6) with respect to θ, which gives

$$\hat\varphi_{i;a}^r = -\hat{\bar l}_i^{rs}\,\hat{\bar l}_{i;as} \tag{8}$$

where $\hat{\bar l}_{i;as} = n_i^{-1}\sum_j w_{ij} l_{j;as}(\theta,\hat\varphi_i)$ and $\hat{\bar l}_i^{rs}$ is the matrix inverse of $\hat{\bar l}_{i;rs} = n_i^{-1}\sum_j w_{ij} l_{j;rs}(\theta,\hat\varphi_i)$. Applying simple expansions to the observed Fisher matrices leads to

$$\hat{\bar l}_{i;as} = -\bar\kappa_{i;a,s} + \bar U_{i;as} + \bar\delta_{i;as} + \bar\kappa_{i;ast}(\hat\varphi_t^t - \varphi_i^t) + \ldots, \tag{9}$$

$$-\hat{\bar l}_i^{\,rs} = \bar\kappa_i^{r,s} + \bar\kappa_i^{r,u}\,\bar\kappa_i^{s,t}\,\{\bar U_{i;tu} + \bar\delta_{i;tu} + \bar\kappa_{i;tuv}(\hat\varphi_i^v - \varphi_i^v)\} + \ldots \tag{10}$$

which in turn finally allows to calculate the bias of the profile score function (4). As shown in the appendix, making use of (7) and inserting (9) and (10) in (5) gives the overall bias

$$\begin{aligned}\sum_i E\{l_{i;a}(\theta,\hat\varphi_i)\} &= \sum_i\{-c_i\,\bar\kappa_i^{r,s}(\bar\kappa_{i;a,rs} - \bar\kappa_{i;a,t}\,\bar\kappa_i^{t,u}\,\bar\kappa_{i;rs,u}) \\ &\qquad -c_i\,\bar\kappa_i^{r,s}(\bar\kappa_{i;a,r,s} - \bar\kappa_{i;a,t}\,\bar\kappa_i^{t,u}\,\bar\kappa_{i;r,s,u}) \\ &\qquad + O(h^4) + O(n_i^{-1}h^2) + O(n_i^{-2})\}\end{aligned} \tag{11}$$

where $c_i = (2n_i - m_i)/(2n_i^2)$ and $m_i = \sum_j w_{ij}^2$. The leading components in (11) are now defined as bias $b_a(\theta)$ so that the adjusted profile score function $l_a^{AP}(\theta) = l_a^P(\theta) - b_a(\theta)$ results. The bias formula also holds if $\hat{\varphi}_{i;a}^r$ in (4) is replaced by its expectation, i.e. if local (estimated) Fisher matrices are used instead of observed Fisher matrices.

Remark 1: The leading component $b_a(\theta)$ has order $\sum_i O(n_i^{-1}) = O(h^{-p})$ and hence tends to infinity for $h \to 0$. This is in contrast to parametric models where the profile score functions typically is biased up to order $O(1)$. The asymptotic correction terms given in (11) are negligible if a) $\sum_i O(h^4)/h^{-p} = nh^{4+p} \to 0$, b) $\sum_i O(n_i^{-1}h^2)/h^{-p} = h^2 \to 0$ and c) $\sum O(n_i^{-2})/h^{-p} = n^{-1}h^{-p} \to 0$. While conditions b) and c) follow by the usual bias-variance trade-off, condition a) is not standard and requires undersmoothing of $\varphi(\cdot)$. The component a) consists of squared smoothing bias terms, e.g. $\sum_i \kappa_{i;a,r}\bar{\kappa}_i^{r,s,t}\bar{\delta}_{i,s}\bar{\delta}_{i,t}$ is a representative. If the order of the smoothing bias is reduced one gets $b_a(\theta)$ as asymptotically dominating term. In practice however, undersmoothing can be a burden since standard data driven bandwidth selection routines optimize the bias-variance trade-off. One should however keep in mind that a bias adjustment by $b_a(\theta)$ can still have a positive effect, even if $\varphi(\cdot)$ is not undersmoothed. This holds since the smoothing bias frequently is small, not in an asymptotic sense but in a practical sense. In a simulation study given in the next section we demonstrate this point.

Remark 2: The structure of formula (11) shows similarities to McCullagh & Tibshirani (1990). In fact if the component $\varphi(\cdot) \equiv \varphi$ is constant one might transform local likelihood to standard likelihood by setting the smoothing parameter $h \to \infty$. This leads to weights $w_{ij} \equiv 1$ and standard likelihood estimates for φ result. The components involved in (11) are then standard likelihood cumulants and $n_i = m_i = n$ such that $c_i = 1/(2n)$. Hence, in this case (11) coincides with the correction given in McCullagh and Tibshirani.

Remark 3: If only the first component of the profile score function (4) is considered, the resulting bias of the profile score function is disturbed by an additional bias component resulting from smoothing, i.e. one has $E\{\sum_i l_{i;a}(\theta, \hat{\varphi}_i)\} = O(h^{-p}) + O(nh^2) + \ldots$ where the $O(nh^2)$ component equals $-\sum_i \kappa_{i;a,r}\,\bar{\kappa}_i^{r,s}\,\bar{\delta}_{i;s}$, see the appendix for details. This implies that the first order smoothing bias is automatically corrected if θ is estimated from the entire profile score function (4). For normally distributed response in a semiparametric regression model this bias reducing effect of the second component in (4) was first demonstrated by Speckman (1988).

3 Semiparametric Regression Models

We demonstrate the use of the bias correction in the semiparametric regression model $E(y|x) = \mu\{z_1\theta + z_2\varphi(u)\}$ with $z = (z_1, z_2)$ where z_1 and z_2 are vectors of functionally independent covariates. To ensure identifiability we include the intercept in z_2. We assume that $y|x \sim \exp\{\vartheta y - \kappa(\vartheta) + g(y)\}$ follows an exponential family distribution and for simplicity we take $\mu(\cdot)$ as natural link, i.e. $\vartheta = z_1\theta + z_2\varphi(u)$. The components of z_1 are indexed by $(z_a, z_b, \ldots)$ and we use $(z_r, z_s, \ldots)$ for z_2. This allows to write the model as $E(y|x) = \mu\{z_a\theta^a + z_r\varphi^r(u)\}$ and one gets $-l_{i;ar} = \kappa_{i,a,r} = z_{i;a}z_{i;r}\kappa_{i;2}$ and $-l_{i;rs} = \kappa_{i,r,s} = z_{i;r}z_{i;s}\kappa_{i;2}$ where $\kappa_{i;2} = \partial^2\kappa(\vartheta)/(\partial\vartheta)^2$. The adjusted profile score function is then obtained by

$$\begin{aligned} l_a^{AP}(\theta) &= \sum_i \left[\tilde{z}_{i;a}\{y_i - h(z_i\theta + \hat{\varphi}_i)\} - b_{i;a}(\theta)\right] \\ &= l_a^P(\theta) - \sum_i b_{i;a}(\theta) \end{aligned} \tag{12}$$

with $\tilde{z}_{i;a} = z_{i;a} - \bar{\kappa}_i^{r,s} z_{i;r} \sum_j w_{ij} z_{j;a} z_{j;s} \kappa_{j;2}$ and bias

$$\begin{aligned} b_{i;a}(\theta) &= -c_i \bar{\kappa}_i^{r,s}\{\sum_j (w_{ij} z_{j;a} z_{j;r} z_{j;s} \kappa_{j;3})/n_i \\ &\qquad - \sum_j (w_{ij} z_{j;a} z_{j;t} \kappa_{j;2}) \bar{\kappa}_i^{t,u} \sum_j (w_{ij} z_{j;r} z_{j;s} z_{j;u} \kappa_{j;3})/n_i^2\} \end{aligned}$$

In applications the component $\varphi(\cdot)$ is frequently univariate, i.e. the model has the form $E(y|x) = \mu\{z_a\theta^a + \varphi(u)\}$. In this case $\tilde{z}_{i;a}$ simplifies to $\tilde{z}_{i;a} = z_{i;a} - \sum_j w_{ij}\kappa_{j;2}\, z_{j;a} / \sum w_{ij} k_{j;2}$ and the bias equals $b_{i;a}(\beta) = -c_i(\sum_j w_{ij}\kappa_{j;3} \quad z_{j;a} - n_i k_{i;3} \tilde{z}_{i;a} \sum_j w_{ij}\kappa_{j;3}) / \sum w_{ij} k_{j;2}$

Simulation Study: We investigate the benefit of the bias correction by the following simulation study. We consider the semiparametric logistic regression model with y as binary response having expectation $E(y|x) = \text{logit}^{-1}\{z\theta + \varphi(u)\}$. We set $\theta = 1/2$ and simulate from $\varphi(u) = -1 + 4(u - 1/2)^2$. The explanatory quantity u is univariate and takes 15 and 25 equidistant points in $[0, 1]$ while z is taken as binary factor. At each point of u we simulate two outcomes of y with different

sample		$\hat{\theta}^{AP}$		$\hat{\theta}^{P}$		
size	design	bias	m.s.e.	bias	m.s.e.	$\frac{\text{m.s.e}(\hat{\theta}^{AP})}{\text{m.s.e}(\hat{\theta}^{P})}$
30	(a)	0.11	0.59	0.18	0.74	0.79
30	(b)	0.06	0.53	0.12	0.65	0.81
50	(a)	0.01	0.46	0.03	0.52	0.88
50	(b)	0.01	0.34	0.02	0.38	0.89

Table 1: Bias and mean squared error (m.s.e.) of profile and adjusted profile estimates.

h	m.s.e.$(\hat{\theta}^{AP})$	m.s.e.$(\hat{\theta}^{P})$	$\frac{\text{m.s.e}(\hat{\theta}^{AP})}{\text{m.s.e}(\hat{\theta}^{P})}$
0.1	0.54	0.76	0.71
0.2	0.55	0.65	0.84
0.3	0.54	0.63	0.85
0.4	0.54	0.60	0.9

Table 2: Mean squared error of profile and adjusted profile estimates for $n = 40$.

design for z, namely (a) z has a balanced design, and (b) we draw z randomly with $P(z = 1) = \text{logit}^{-1}(-1.5 + 3u)$. The total sample size is therefore 30, 50 respectively. In each setting we draw 200 simulations to assess the properties of the adjusted estimates. Moreover in each of this simulations we choose the bandwidth h by the Akaike criteria

$$\max\, l^P(\hat{\theta}^P) - \sum_i n_i^{-1}.$$

The penalty term $\sum_i n_i^{-1}$ is frequently called the degree of freedom for a smooth model (see e.g. Hastie & Tibshirani, 1990). The results are reported in Table 1. The adjusted estimate clearly shows a reduced bias. Moreover the mean squared error is reduced due to the adjustment. This effects shows for fixed design as well as for random design.

We run a second simulation study with $n = 40$ and fixed design to investigate the effect of the bandwidth separately. The selection routine chosen above is not optimal in the sense that $\varphi(\cdot)$ is not undersmoothed, see Remark 1 above. Table 2 shows the mean squared error of the estimators for various fixed bandwidths. It appears that the adjusted estimate possesses a rather stable mean squared error while the mean squared error of $\hat{\theta}^P$ decreases for large bandwidth. A similar behavior was observed for other simulations which are not reported here.

4 Discussion

We showed above that the bias of the profile score function in semiparametric models is of order $O(h^{-p})$ with h as bandwidth. The bias can be adjusted using techniques similar to those used in parametric models. In simulations we demonstrated that the adjustment of the bias can also improve the mean squared error. As can be seen from the appendix, the expansions required for the calculation of the bias are more complicated compared to those found in a solely parametric framework. This has two reasons, first the profile score function (4) consists of two components, while the second component in (4) is zero in parametric models. Second, smooth estimates are biased and therefore do not possess a $\sqrt{n}$ convergence. Though it would in principal be possible to also adjust the information bias by extending the expansions above to the second order derivative, the complicated structure of the formulae in the semiparametric framework makes it rather awkward to correct the information bias analytically. Instead, one can pursue a numerical approach as proposed by McCullagh & Tibshirani (1990) for parametric models. We give a short sketch of the procedure in the appendix.

A Technical Details

We first show that smoothing bias components denoted by δ have order $O(h^2)\{1 + O_p(n_i^{-1/2})\}$. One easily finds

$$\begin{aligned}\bar{\delta}_{i;r} &= n_i^{-1}\sum_j w_{ij} l_{j;rs}(\varphi_i^s - \varphi_j^s) + n_i^{-1}\sum_j w_{ij} l_{j;rst}(\varphi_i^s - \varphi_j^s)(\varphi(u_i)^t - \varphi_j^t) + \dots \\ &= -n_i^{-1}\sum_j w_{ij}\kappa_{j;r,s}(\varphi_i^s - \varphi_j^s) + n_i^{-1}\sum_j w_{ij} U_{j;rs}(\varphi(u_i)^s - \varphi_j^s) + \dots \end{aligned} \tag{13}$$

To see that the first component in (13) is $O(h^2)$ let $\bar{\kappa}_{u;r,s} = -E_z\{E(l_{i;rs})|u\}$ denote the mean cumulant where the inner expectation is taken with respect to density $f(y|x;\theta,\varphi(u))$ and the outer expectation uses the design density $f(x|u)$. This allows by standard kernel smoothing arguments to obtain

$$n_i^{-1}\sum_j w_{ij}\kappa_{j;r,s}(\varphi_i^s - \varphi_j^s) \approx h^{-p}\int K\left(\frac{u_i - u}{h}\right)\bar{\kappa}_{u;r,s}(\varphi_i^s - \varphi^s(u))f(u)du$$

$$
\begin{aligned}
&= \bar{\kappa}_{u_i;r,s} h^{-p} \int K\left(\frac{u_i - u}{h}\right)(\varphi_i^s - \varphi^s(u)) f(u) du + \ldots \\
&= O(h^2)
\end{aligned}
$$

In the same fashion one finds that the the second component in (13) has order $O_p(n_i^{-1/2}h^2)$ and hence can generally be neglected.

Series inversions (see e.g. Barndorff-Nielsen & Cox, 1989) directly allows to solve (6) about $\hat{\varphi}_i^r - \varphi_i^r$ which gives (7). To gain more insight in the structure of asymptotic correction terms listed in (7) we give a representative for each of the terms:

$$
\begin{aligned}
O(h^6) &= a^{rst}\bar{\delta}_{i,r}\bar{\delta}_{i,s}\bar{\delta}_{i,t} + \ldots \\
O_p(n_i^{-1/2}h^4) &= b^{rstu}\bar{\delta}_{i,rs}\bar{\delta}_{i,t}\bar{U}_{i;u} + \ldots \\
O_p(n_i^{-1}h^2) &= c^{rstu}\bar{\delta}_{i,rs}\bar{U}_{i;t}\bar{U}_{i;u} + \ldots \\
O_p(n_i^{-3/2}) &= c^{rstu}\bar{U}_{i;rs}\bar{U}_{i;t}\bar{U}_{i;u} + \ldots
\end{aligned}
$$

with a^{rst}, b^{rstu}, c^{rstu} and d^{rstu} are some arrays with elements of order $O(1)$.

Let us now consider the first component in (4). Expansion yields

$$
\sum_i \hat{l}_{i;a} = \sum_i l_{i;a}(\theta, \hat{\varphi}_i) = \sum_i \left\{ l_{i;a} + l_{i;ar}(\hat{\varphi}_i^r - \varphi_i^r) + \frac{1}{2} l_{i;ars}(\hat{\varphi}_i^r - \varphi_i^r)(\hat{\varphi}_i^s - \varphi_i^s) \right\} + \ldots . \tag{14}
$$

and inserting (7) gives

$$
\begin{aligned}
\sum_i \hat{l}_{i;a} \quad = \quad & \sum_i \left[l_{i;a} + l_{i;ar}\{\bar{\kappa}_i^{r,s}(\bar{U}_{i;s} + \bar{\delta}_{i;s}) + \bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\bar{U}_{i;st}\,\bar{U}_{i;u} + \frac{1}{2}\bar{\kappa}_i^{r,s,t}\,\bar{U}_{i;s}\,\bar{U}_{i;t}\} \right. \\
& \left. + \frac{1}{2} l_{i;ars}\,\bar{\kappa}_i^{r,t}\,\bar{\kappa}_i^{s,u}\,\bar{U}_{i;t}\,\bar{U}_{i;u} \right] + \ldots .
\end{aligned}
$$

where $\bar{\kappa}_i^{r,s,t} = \bar{\kappa}_i^{r,u}\,\bar{\kappa}_i^{s,v}\,\bar{\kappa}_i^{t,w}\,\bar{\kappa}_{i;uvw}$. Taking expectation leads to

$$
\begin{aligned}
\sum_i E\{l_{a;i}(\theta, \hat{\varphi}_i)\} \quad = \quad & \sum_i n_i^{-1} \Big\{ \bar{\kappa}_i^{r,s}\,\kappa_{i;ar,s} - n_i^{-1}\kappa_{i;a,r}\,\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u} \sum_j w_{ij}^2 \kappa_{j;st,u} \\
& - \frac{1}{2} n_i^{-1}\kappa_{i;a,r}\,\bar{\kappa}_i^{r,s,t} \sum_j w_{ij}^2 \kappa_{j;s,t} \\
& + \frac{1}{2} n_i^{-1}\kappa_{i;ars}\,\bar{\kappa}_i^{r,t}\,\bar{\kappa}_i^{s,u} \sum_j w_{ij}^2 \kappa_{j;t,u} - \kappa_{i;a,r}\bar{\kappa}_i^{r,s}\bar{\delta}_{i;s} \\
& + O(h^4) + O(n_i^{-1}h^2) + O(n_i^{-2}) \Big\}.
\end{aligned} \tag{15}
$$

To clarify the structure of the asymptotic correction terms we again list a representative for each of the components:

$$
\begin{aligned}
\sum_i O(h^4) &= \sum_i a_i^{rs}\bar{\delta}_{i;r}\bar{\delta}_{i;s} + \dots \\
\sum_i O(n_i^{-1}h^2) &= \sum_i n_i^{-1} b_i^r \bar{\delta}_{i;r} + \dots \\
\sum_i O(n_i^{-2}) &= \sum_i n_i^{-2} c_i + \dots
\end{aligned}
$$

with a_i^{rs}, b_i^r and c_i denoting arrays of order $O(1)$ here.

Formula (15) can be simplified by reflecting that symmetrical kernels fulfill

$$
m_i^{-1}\sum_j w_{ij}^2 \kappa_{j;t,u} \;=\; n_i^{-1}\sum_j w_{ij}\kappa_{j;t,u} + O(h^2) \;=\; \bar{\kappa}_{i;t,u} + O(h^2)
$$

with $m_i = \sum_j w_{ij}^2$. Moreover, Bartlett's third order identity $\kappa_{i;ars} = -\kappa_{i;ar,s} - \kappa_{i;as,r} - \kappa_{i;a,rs} - \kappa_{i;a,r,s}$ allows to get from (15)

$$
\begin{aligned}
\sum_i E\{l_{i;a}(\theta,\hat{\varphi}_i)\} \;=\; \sum_i \Big\{ &-\frac{m_i}{2n_i^2}\,\bar{\kappa}_i^{r,s}(\kappa_{i;a,rs} - \kappa_{i;a,t}\,\bar{\kappa}_i^{t,u}\,\bar{\kappa}_{i;u,rs}) \\
&-\frac{m_i}{2n_i^2}\,\bar{\kappa}_i^{r,s}(\kappa_{i;a,r,s} - \kappa_{i;a,t}\,\bar{\kappa}_i^{t,u}\,\bar{\kappa}_{i;r,s,u}) \\
&+\frac{(n_i - m_i)}{n_i^2}\,\bar{\kappa}_i^{r,s}\,\kappa_{i;ar,s} - \kappa_{i;a,r}\,\bar{\kappa}_i^{r,s}\,\bar{\delta}_{i;s} \\
&+O(h^4) + O(n_i^{-1}h^2) + O(n_i^{-2})\Big\}.
\end{aligned}
\tag{16}
$$

In the next step we consider the second component in (4). Expansion allows to write

$$
\begin{aligned}
\sum_i \hat{\varphi}_{i;a}^r l_{i;r}(\theta,\hat{\varphi}_i) \;=\; \sum_i \hat{\varphi}_{i;a}^r\{\, &l_{i;r} + \,l_{i;rt}(\hat{\varphi}_i^t - \varphi_i^t) \\
&+\frac{1}{2}\,l_{i;rtu}(\hat{\varphi}_i^t - \varphi_i^t)(\hat{\varphi}_i^u - \varphi_i^u)\} + \dots.
\end{aligned}
\tag{17}
$$

where $\hat{\varphi}_{i;a}^r = -\hat{\bar{l}}_i^{rs}\hat{\bar{l}}_{i;as}$ can be expanded by making use of

$$
\hat{\bar{l}}_{i;as} \;=\; -\bar{\kappa}_{i;a,s} + \bar{U}_{i;as} + \bar{\delta}_{i;as} + \bar{\kappa}_{i;ast}(\hat{\varphi}_i^t - \varphi_i^t) + \dots \tag{18}
$$

$$
-\hat{\bar{l}}_i^{rs} \;=\; \bar{\kappa}_i^{r,s} + \bar{\kappa}_i^{r,u}\,\bar{\kappa}_i^{s,t}\,\{\bar{U}_{i;tu} + \bar{\delta}_{i;tu} + \bar{\kappa}_{i;tuv}(\hat{\varphi}_i^v - \varphi_i^v)\}. \tag{19}
$$

Inserting these terms and expansion (7) in (17) gives

$$
\begin{aligned}
&\sum_i \hat{\varphi}_{i;a}^r l_{i;r}(\theta,\hat{\varphi}_i) \\
=\; &\sum_i -\bar{\kappa}_i^{r,s}\bar{\kappa}_{i;a,s}\big[l_{i;r} + l_{i;rt}\{\bar{\kappa}_i^{t,u}(\bar{U}_{i;u} + \bar{\delta}_{i;u}) + \bar{\kappa}_i^{t,u}\bar{\kappa}_i^{v,w}\bar{U}_{i;uv}\bar{U}_{i;w} + \frac{1}{2}\bar{\kappa}^{t,u,v}\bar{U}_{i;u}\bar{U}_{i;v}\}
\end{aligned}
$$

$$+\frac{1}{2}l_{i;rtu}\bar{\kappa}_i^{t,v}\bar{\kappa}_i^{u,w}\bar{U}_{i;v}\bar{U}_{i;w}]$$
$$+\bar{\kappa}_i^{r,s}(\bar{U}_{i;as}+\bar{\kappa}_{i;ast}\,\bar{\kappa}_i^{t,u}\,\bar{U}_{i;u})(l_{i;r}+l_{i;rv}\bar{\kappa}_i^{v,w}\bar{U}_{i;w})$$
$$-\bar{\kappa}_i^{r,u}\bar{\kappa}_i^{s,t}(\bar{U}_{i;tu}+\bar{\kappa}_{i;tuv}\bar{\kappa}_i^{v,w}\bar{U}_{i;w})\bar{\kappa}_{i;a,s}(l_{i;r}+l_{i;rx}\bar{\kappa}_i^{x,y}\bar{U}_{i;y})+\ldots.$$

Taking expectation then permits

$$\begin{aligned}\sum_i E(\hat{l}_{r;i}\hat{\varphi}^r_{i;a}) \;=\; & \sum_i n_i^{-1}\Big\{-\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\kappa_{i;rt,u}\,\bar{\kappa}_{i;a,s} \\ & +\frac{m_i}{n_i}\bar{\kappa}_i^{r,s}\,\kappa_{i;r,t}\,\bar{\kappa}_{i;a,s}\big(\bar{\kappa}_i^{t,u}\,\bar{\kappa}_i^{v,w}\,\bar{\kappa}_{i;uv,w}+\frac{1}{2}\bar{\kappa}_i^{t,u,v}\,\bar{\kappa}_{i;u,v}\big) \\ & -\frac{m_i}{2n_i}\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\kappa_{i;rtu}\bar{\kappa}_{i;a,s} \\ & +\bar{\kappa}_i^{r,s}\,\kappa_{i;as,r}+\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\kappa_{i;r,u}\,\bar{\kappa}_{i;ast}-\frac{m_i}{n_i}\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\kappa_{i;r,t}\big(\bar{\kappa}_{i;as,u}+\bar{\kappa}_{i;asu}\big) \\ & -\bar{\kappa}_i^{r,u}\,\bar{\kappa}_i^{s,t}\,\kappa_{i;r,tu}\,\bar{\kappa}_{i;a,s}-\bar{\kappa}_i^{r,s,t}\,\kappa_{i;r,t}\,\bar{\kappa}_{i;a,s} \\ & +\frac{m_i}{n_i}\bar{\kappa}_i^{r,u}\,\bar{\kappa}_i^{s,v}\,\bar{\kappa}_i^{t,w}\,\kappa_{i;r,t}\,\bar{\kappa}_{i;a,s}\big(\bar{\kappa}_{i;uv,w}+\bar{\kappa}_{i;uvw}\big) \\ & +\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\kappa_{i;r,t}\bar{\kappa}_{i;a,s}\,\bar{\delta}_{i;u}+O(h^4)+O(n_i^{-1}h^2)+O(n_i^{-2})\Big\}.\end{aligned} \tag{20}$$

In order to simplify the formula above we can use approximations of the type $\sum_i \bar{\kappa}_i^{r,s}\kappa_{i;r,t}\bar{\kappa}_{i;a,s} = \sum_i \bar{\kappa}_i^{r,s}\bar{\kappa}_{i;r,t}\bar{\kappa}_{i;a,s}$, i.e. the distinction between $\kappa_{i;r,t}$ and $\bar{\kappa}_{i;r,t}$ is not required when summing over i. To see this we employ the notation $\bar{\kappa}_{u_i;r,t} = -E_z\{E(l_{i;rt})|u_i\}$ from above to denote the mean second order derivative where the expectation is taken with respect to $f(y|x;\theta,\varphi(u_i)$ and the design density $f(x|u_i)$. This notation allows to write

$$\sum_i \bar{\kappa}_i^{r,s}\kappa_{i;r,t}\bar{\kappa}_{i;a,s} \;\approx\; n\int \bar{\kappa}_u^{r,s}(\int \kappa_{u;r,t}f(x|u)dx)\bar{\kappa}_{u,a,s}f(u)du \approx \sum_i \bar{\kappa}_i^{r,s}\bar{\kappa}_{i;r,t}\bar{\kappa}_{i;a,s}.$$

Making use of such approximations allows to simplify (20) and one gets

$$\begin{aligned}\sum_i E(\hat{l}_{i;r}\hat{\varphi}^r_{i;a}) \;=\; & \sum_i \frac{n_i-m_i}{n_i^2}\bar{\kappa}_i^{r,s}\,\bar{\kappa}_i^{t,u}\,\bar{\kappa}_{i;a,s}(\bar{\kappa}_{i;r,tu}+\bar{\kappa}_{i;r,t,u}) \\ & -\frac{n_i-m_i}{n_i^2}\bar{\kappa}_i^{r,s}(\bar{\kappa}_{i;ar,s}+\bar{\kappa}_{i;a,rs}+\bar{\kappa}_{i;a,r,s})+\bar{\kappa}_{i;a,r}\,\bar{\kappa}_i^{r,s}\,\bar{\delta}_{i;s}.\end{aligned} \tag{21}$$

Adding (16) and (21) finally gives (11). If the observed Fisher matrices (18) and (19) are replaced by estimated versions we have to calculate the expectation of $E(-\hat{l}_{i;r}\hat{\bar{\kappa}}_i{}^{rs}\hat{\bar{\kappa}}_{i;sa})$. The derivation is similar, except that (18) and (19) turn into

$$\begin{aligned}\hat{\bar{\kappa}}_{i;as} &= -\bar{\kappa}_{i;a,s}+(\bar{\kappa}_{i;ast}+\bar{\kappa}_{i;as,t})(\hat{\varphi}_i^t-\hat{\varphi}_i^t) \\ \hat{\bar{\kappa}}_i{}^{rs} &= -\bar{\kappa}_i^{r,s}-\bar{\kappa}_i^{r,u}\,\bar{\kappa}_i^{s,t}(\bar{\kappa}_{i;tuv}+\bar{\kappa}_{i;tu,v})(\hat{\varphi}_i^v-\varphi_i^v).\end{aligned}$$

Calculation in the above fashion again proves the validity of (11)

Sketch of information bias adjustment

Assume for simplicity that θ is univariate. Calculation of $l_a^{AP}(\theta)$ for a grid of points easlily allows to obtain the second order derivative $l_{aa}^{AP}(\theta)$ by numerical differentiation. Moreover one can use bootstrapping to estimate $E\{l_a^{AP}(\theta)l_a^{AP}(\theta)\}$. Let therefore y_i^* be drawn from $f(y|x_i, \theta, \hat{\varphi}(u_i))$ for $i = 1, \dots n$, and let $\hat{\varphi}_i^*$ be the smooth fit obtained from (3) by replacing y_i with y_i^*. This yields $l_a^{*^{AP}}(\theta)$ as bootstrapped profile score function. Drawing now B bootstrap samples allows to calculate $w_a(\theta) = l_{aa}^{AP}(\theta)/\{\sum_{b=1}^{B} l_a^{*^{AP}}(\theta)l_a^{*^{AP}}(\theta)/B\}$ where the sum is taken over the B bootstrap samples $l_a^{*^{AP}}(\theta)$. Defining now $\tilde{l}_a^{AP}(\theta) = w_a(\theta)l_a^{AP}(\theta)$ then provides $E(\tilde{l}_{aa}^{AP}(\theta)) \approx E(\tilde{l}_a^{AP}\tilde{l}_a^{AP})$.

References

Barndorff-Nielsen, O. E. (1991). Likelihood theory. In D. V. Hinkley, N. Reid, & E. J. Snell (Eds.), *Statistical Theory and Modelling*. London: Chapman and Hall.

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

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. B 49*, 1– 39.

Davison, A. (1988). Aproximate conditional inference in generalized linear models. *J. Roy. Statist. Soc. 50*, 445–461.

DiCiccio, T. J., Martin, M. A., Stern, S. E., and Young, G. A. (1996). Information bias and adjusted profile likelihoods. *J. Roy. Statist. Soc. B 58*, 189–203.

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

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.

McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman & Hall.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* (second ed.). New York: Chapman and Hall.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. B 52*, 325–344.

Reid, N. (1995). The role of conditioning in inference. *Statist. Science 10*, 138–199.

Severini, T. A. (1998). *An Approximation to the modified profile likelihood function*, Volume 85. Biometrika.

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

Severini, T. A. and 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. (1989). The kernel estimate of a regression function in likelihood–based models. *J. Amer. Statist. Assoc. 84*, 276–283.

Stern, S. E. (1997). A second-order adjustment of the profile likelihood in the case of a multidimensional parameter of interests. *J. Roy. Statist. Soc. B, 59*, 653–665.