Quasi-Likelihood Regression with Multiple Indices and Smooth Link and Variance Functions

October 2003 (Revised Version)

JENG-MIN CHIOU

National Health Research Institutes, Taiwan

HANS-GEORG MÜLLER

University of California, Davis

Running headline: Multiple-index quasi-likelihood regression

Correspondence: Jeng-Min Chiou, *Email*: jmchiou@nhri.org.tw, *Phone*: 886-2-26534401 ext.7630, *Fax*: 886-2-27890253. *Address*: Division of Biostatistics and Bioinformatics, National Health Research Institutes, 128 Yen-Chiu-Yuan Rd. Sec. 2, Taipei 115, Taiwan, R.O.C.

ABSTRACT. A flexible semi-parametric regression model is proposed for modeling the relationship between a response and multivariate predictor variables. The proposed multiple-index model with unknown link and variance functions is an extension of the single index model of Chiou & Müller (1998). The unknown functions are assumed to be smooth and are estimated nonparametrically. We propose data-adaptive methods for automatic smoothing parameter selection and for the choice of the number of indices M. This model is very flexible, easy to use and adapts to complex data structures. It provides efficient adaptive estimation through the variance function component in the sense that the asymptotic distribution is the same as if the nonparametric components are known. We develop iterative estimation schemes which include a constrained projection method for the case where the regression parameter vectors are mutually orthogonal. The proposed methods are illustrated with the analysis of data from a food folate experiment in a rat growth bioassay, and from a medfly reproduction experiment with various feeding schedules. Asymptotic properties of the estimated model components are also obtained.

Key words: constrained estimation, generalized linear models, nonparametric quasi-likelihood, projection pursuit regression, pseudo-likelihood, semi-parametric regression

1. Introduction

Generalized linear models are being widely used as a standard tool in modern regression analysis. Successful modeling based on generalized linear models relies on correctly specified model components including the random part and the systematic part. In a classical generalized linear model, the random part requires specification of a distribution from the exponential family. This distribution assumption can be relaxed through the specification

of a variance function by Wedderburn's (1974) quasi-likelihood approach. The systematic part of the model includes a linear predictor and a link function. The linear predictor typically is a single index, i.e., a linear combination of the predictors. A single index provides a dimension reduction step. The value of the single index is related to the mean through the link function. Correct specification of link and variance functions are key ingredients for successful statistical modeling of a generalized linear model with quasi-likelihood, in the simple situation where a single linear predictor is indeed sufficient for modeling the relationship between covariates and means. We note that consistency of the regression parameter estimates depends on a correctly specified link function, while efficiency depends on a correctly specified variance function. The importance of choosing a correct link function has been noted in the literature; see, e.g. Zhang (1999, 2000, 2001). On the other hand, the price of mis-specifying the variance function is not only loss of efficiency of the regression parameter estimates but also incorrect confidence regions and test results.

Chiou & Müller (1998) proposed a modified quasi-likelihood regression method where link and variance functions are unknown and are estimated nonparametrically. They obtained consistency results for the link and the variance function estimates as well as the asymptotic distribution of the regression coefficients. If the single index assumption is correct, this QLUE (Quasi-Likelihood with Unknown link and variance function Estimation) model requires nothing more from the user than specification of the predictor variables, and thus is as easy to use as a classical multiple linear regression model. However, as the successful modeling for increasingly large and complex data requires even more flexible models, the single index assumption may prove to be too restrictive for some applications.

While the choice of a suitable link function allows to include nonlinear relationships with a single index, a single index is still a limiting factor which may lead to inadequate results. One approach to increase flexibility is to include nonlinear components in the

linear predictor. For instance, Carroll, Fan, Gijbels & Wand (1997) proposed generalized partially linear single-index models by adding a nonparametric component to the single linear predictor for the case of known link and variance functions. Hastie & Tibshirani (1986, 1990) proposed a generalized additive model which is different from the conventional generalized linear model in that the linear predictor is replaced with additive predictors, each coupled with an arbitrary univariate function. However, the assumption of additive action of various predictors can be limiting, and this method is not particularly suitable for discrete predictor variables, which are common in applications. Another approach to move beyond a single index is projection pursuit regression (PPR). A standard algorithm for PPR was first proposed by Friedman & Stuetzle (1981), and Roosen & Hastie (1994) provided an alternative algorithm for PPR where the smooth functions are estimated using smoothing splines. More recently, Lingjærde & Liestøl (1998) proposed the generalized projection pursuit regression model which is an extension of PPR to exponential family distributions and allows multiple responses and nonlinear projections of the variables. In practice, it is not easy to justify the assumption on the exponential family distribution under complex situations, especially for the variance function. Moreover, asymptotic distributions for the estimates of model components remain to be developed for statistical inference.

In this study, we propose a MUltiple-index SEmi-parametric quasi-likelihood (MUSE) model that includes an unknown nonparametric variance function, allowing dependence of the variance on the means with an unknown form. This is an extension of and repeatedly makes use of the single index quasi-likelihood regression model with unknown link and variance functions (QLUE model) of Chiou & Müller (1998). The proposed model extends the traditional single linear predictor to the case of multiple linear predictors or indices, each coupled with an unknown link function. The multiple linear predictors are assumed to act in an additive manner. The method of constrained projection is employed to obtain

the parameter estimates for the multiple linear predictors, where the projection indices are forced to be orthogonal to one another.

This article contains not only new methodology with theoretical results on identifiability and asymptotic distributions, but also several data analyses and discussion of practical issues, thus demonstrating both theoretical and practical facets of the proposed methods. It is organized as follows. The proposed model and assumptions are described in section 2, including a discussion of identifiability. Details of the estimation procedure for the model components are provided in section 3. The method of constrained estimation of the regression parameters is the theme of section 4. Section 5 contains data analyses regarding the effect of folates on growth rates and the effect of various feeding schemes on medfly reproduction. Asymptotic properties for the estimated model components are presented in section 6. Concluding remarks are in section 7, while details and proofs are compiled in the appendices.

2. The proposed MUSE model

2.1. Model and assumptions

The proposed multiple-index semi-parametric quasi-likelihood (MUSE) model includes M indices which are linear combinations of the predictor variables. We assume there are n independent observations y_i of a response variable with associated p-variate predictors $\boldsymbol{x}_i$, $p \geq 1$, and M link functions $g_k(\cdot)$ and regression parameter vectors $\boldsymbol{\beta}_k$, $k = 1, \ldots, M$, $1 \leq M \leq p$, such that the observations y_i and the p-dimensional predictors $\boldsymbol{x}_i$ are related by

$$E(y_i) = \sum_{k=1}^{M} g_k(\boldsymbol{x}_i^T \boldsymbol{\beta}_k). \tag{1}$$

When $M = 1$, this is a single-index model, corresponding to the QLUE model proposed by Chiou & Müller (1998). A second important assumption is that there exists a variance

function $\sigma^2(\cdot)$, $\sigma^2(\cdot) \geq \gamma > 0$, such that

$$Var(y_i) = \sigma^2(E(y_i)). \tag{2}$$

The variance of the observations is assumed to be solely a function of the means; this function is referred to as the variance function, as is customary in generalized linear models. We assume that the link and variance functions $g_k(\cdot)$, $k = 1, \ldots, M$, and $\sigma^2(\cdot)$ are unknown but smooth. More precisely, we require the following technical assumptions.

(M1) The link functions $g_k(\cdot)$, $k = 1, \ldots, M$ are three times and the variance function $\sigma^2(\cdot)$ is twice continuously differentiable with bounded derivatives. Each link function is strictly monotone.

(M2) There exists a function $\mu_4(\cdot)$ such that $E(\varepsilon^4) = \mu_4(E(y_i))$, where $\varepsilon = y_i - E(y_i)$. The function $\mu_4(\cdot)$ is continuous and there exists an $s \geq 4$ such that $\max\limits_{1 \leq i \leq n} E(\varepsilon^{2s}) < c < \infty$ for some constant $c > 0$.

Additional requirements for data and designs are:

(M3) There exists a $R > 0$ such that $\max\limits_{1 \leq i \leq n} \|\boldsymbol{x}_i\| \leq R < \infty$, for all n.

We assume that $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ form a sequence of designs such that for each k, the linear predictors $\eta_{k_i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_k$ are generated by a "design density" f_{η_k} which is assumed to satisfy the following conditions:

(M4) The support of f_{η_k}, $k = 1, \ldots, M$, is a compact interval, $\int f_{\eta_k}(t)dt = 1$, and f_{η_k} is twice continuously differentiable, satisfying $0 < \inf f_{\eta_k}(\cdot) \leq \sup f_{\eta_k}(\cdot) < \infty$. The design points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ are chosen in such a way that the values $\eta_{k_i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_k$ satisfy

$$\int_{-\infty}^{\eta_{k_i}} f_{\eta_k}(t)dt = \frac{i-1}{n-1} \quad \text{for all } n.$$

(M5) There exist positive definite matrices Σ_k, $k = 1, \ldots, M$, such that, as $n \to \infty$,

$$\frac{1}{n}({D_k}^T V^{-1} D_k) \to \Sigma_k,$$

where D_k is the $n \times p$ matrix of full rank with elements $(D_k)_{ir} = {g_k}'(\eta_{k_i}) x_{ir}$, $\eta_{k_i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_k$, for $1 \leq i \leq n$ and $1 \leq r \leq p$, and V is a diagonal matrix with elements $\{\sigma^2(\mu_{k_i})\}$ for $1 \leq i \leq n$.

We note that (M1) and (M4) imply that there exists a design density f_μ for the means $\mu_{k_i} = g_k(\eta_{k_i})$. For technical assumptions on the kernel function and the smoothing parameters of the QLUE approach, we refer to assumptions (K1)-(K3) in Chiou & Müller (1998). We refer to such conditions as regularity conditions for smoothing. We further note that when $M = 1$ with correctly specified link and variance functions, McCullagh (1983) showed that the quasi-likelihood estimate $\hat{\boldsymbol{\beta}}_1$ of the regression parameters $\boldsymbol{\beta}_1$ is then asymptotically normally distributed under mild regularity conditions,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma_1^{-1}).$$

2.2. Identifiability

Additional assumptions are required to ensure that our proposed model is identifiable. The following condition is sufficient.

(M6) The parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M$ are linearly independent p-vectors, $1 \leq p \leq M$, and satisfy $\|\boldsymbol{\beta}_k\| = 1$, where $\|\cdot\|$ denotes the Euclidean norm. Furthermore, for any linearly independent set of vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M \in \mathbb{R}^p$, such that $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i$ does not hold for all $i = 1, \ldots, M$, for $M \geq 2$, there exists a $c \in \mathbb{R}$ such that the level set $L_c = \{\boldsymbol{x} \in \mathbb{R}^p | \sum_{k=1}^{M} g_k(\boldsymbol{x}^T \boldsymbol{\beta}_k) = c\}$ has the following property: There exist $\boldsymbol{x}_1, \boldsymbol{x}_2 \in L_c$ and $i_0 \in \{1, \ldots, M\}$ such that $\boldsymbol{x}_1^T \boldsymbol{\alpha}_i = \boldsymbol{x}_2^T \boldsymbol{\alpha}_i$, $i \neq i_0$ for $1 \leq i \leq M$, and $\boldsymbol{x}_1^T \boldsymbol{\alpha}_{i_0} \neq \boldsymbol{x}_2^T \boldsymbol{\alpha}_{i_0}$.

We note that the parameter vectors do not contain an intercept, as this would not be identifiable in view of the nonparametric link functions $g_k(\cdot)$, $k = 1, \ldots, M$. Furthermore, condition (M6) implies identifiability of the multiple-index model. The proof is in Appendix A.

As an illustration of the identifiability condition (M6), consider the case where $p = 2$, $M = 2$ and $g_1(u) = u^2$, $g_2(u) = u$. In this case, for a constant c, $L_c = \{\boldsymbol{x}^T = (x_1, x_2)^T \in \mathbb{R}^2 | (\boldsymbol{x}^T\boldsymbol{\beta}_1)^2 + (\boldsymbol{x}^T\boldsymbol{\beta}_2) = c\}$. Setting $\boldsymbol{\alpha}_1 = (\alpha_{11}, \alpha_{12})^T$, assume without loss of generality that $\alpha_{11} \neq 0$, and consider the additional constraint,

$$\boldsymbol{x}^T\boldsymbol{\alpha}_1 = c_1, \tag{3}$$

for a constant c_1. We can substitute x_1 by x_2 from (3). Setting $\tilde{a} = (\beta_{12} - \alpha_{12}\beta_{11}/\alpha_{11})^2$, $\tilde{b} = 2(\beta_{22} - \alpha_{12}\beta_{21}/\alpha_{11})(\beta_{12} - \alpha_{12}\beta_{11}/\alpha_{11})(\beta_{11}c_1/\alpha_{11})$ and $\tilde{c} = (\beta_{11}c/\alpha_{11})^2 + \beta_{21}c/\alpha_{11} - c$, the resulting equation for x_2 is

$$\tilde{a}x_2^2 + \tilde{b}x_2 + \tilde{c} = 0. \tag{4}$$

The assumptions that $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ form a different basis from $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\|\boldsymbol{\alpha}_i\| = \|\boldsymbol{\beta}_i\|$, $i = 1, 2$, immediately imply that $\tilde{a} \neq 0$. By calculation we find that equation (4) has two roots if and only if $c > \beta_{21}(c_1 + \alpha_{12})/\alpha_{11} - \beta_{22}$. Therefore, we always can find c, c_1 such that two distinct roots $\tilde{\boldsymbol{x}}_1$, $\tilde{\boldsymbol{x}}_2$ exist. We then have by construction $\tilde{\boldsymbol{x}}_1^T\boldsymbol{\alpha}_1 = \tilde{\boldsymbol{x}}_2^T\boldsymbol{\alpha}_1$, and $\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2 \in L_c$. At the same time, we must have $\tilde{\boldsymbol{x}}_1^T\boldsymbol{\alpha}_2 \neq \tilde{\boldsymbol{x}}_2^T\boldsymbol{\alpha}_2$ since otherwise it would follow that $\tilde{\boldsymbol{x}}_1 = \tilde{\boldsymbol{x}}_2$. In this way, we have constructed points $\tilde{\boldsymbol{x}}_1$, $\tilde{\boldsymbol{x}}_2$ for which

$$g_1(\tilde{\boldsymbol{x}}_1^T\boldsymbol{\beta}_1) + g_2(\tilde{\boldsymbol{x}}_1^T\boldsymbol{\beta}_2) = g_1(\tilde{\boldsymbol{x}}_2^T\boldsymbol{\beta}_1) + g_2(\tilde{\boldsymbol{x}}_2^T\boldsymbol{\beta}_2),$$

but

$$\tilde{g}_1(\tilde{\boldsymbol{x}}_1^T\boldsymbol{\alpha}_1) + \tilde{g}_2(\tilde{\boldsymbol{x}}_1^T\boldsymbol{\alpha}_2) \neq \tilde{g}_1(\tilde{\boldsymbol{x}}_2^T\boldsymbol{\alpha}_1) + \tilde{g}_2(\tilde{\boldsymbol{x}}_2^T\boldsymbol{\alpha}_2).$$

This means that the MUSE model is identifiable in this example, illustrating a special case of the identifiability condition (M6).

3. Estimating the model components

The estimation of the model components, $\boldsymbol{\beta}_k$, $g_k(\cdot)$ and $\sigma^2(\cdot)$, are based on the QLUE approach which serves as a building block. Starting values for estimating the model component may be obtained by first estimating the regression coefficients $\boldsymbol{\beta}_k$ via sliced inverse regression (SIR) by Li (1991). We start with a brief review of the basic QLUE approach for the case where $M = 1$.

3.1. Quasi-likelihood with unknown link and variance functions

Given the observations y_i and the predictors $\boldsymbol{x}_i$, the first two moments of the response variable are then given by $E(y_i) = g_1(\boldsymbol{x}_i^T\boldsymbol{\beta}_1)$, $\mathrm{var}(y_i) = \mathrm{var}(\varepsilon_i) = \sigma^2(E(y_i))$ for $i = 1, \ldots, n$. The estimation procedures for this model include estimation of nonparametric components, namely link function $g_1(\cdot)$ and variance function $\sigma^2(\cdot)$, and of the parametric component, the regression parameter vector $\boldsymbol{\beta}_1$. To simplify the notation, let $S^{(\nu)}(w, b; (w_i, y_i)_{i=1,\ldots,n})$, $\nu \geq 0$, be a generic notation for a nonparametric estimator or a one-dimensional smoothing method, targeting the νth derivative of a function, $d^\nu/dw^\nu E(Y|W = w)$, based on scatterplot data $(w_i, y_i)_{i=1,\ldots,n}$. Here the w_i's are design points, y_i's are the raw measurements to be smoothed, b denotes a bandwidth or a smoothing parameter, and w is a target point at which the function is evaluated. Let $g_1^{(\nu)}$ denote the νth derivative of the link function g_1.

In the QLUE three-stage iterative estimation procedure, the parametric and nonparametric estimation steps are alternated. The procedure can be summarized as follows (for further details see Chiou & Müller, 1998).

1. (*Nonparametric estimation step for link function*) Given $\hat{\boldsymbol{\beta}}_1$, the estimates of the link function and its first derivative $g_1^{(\nu)}$, $\nu = 0, 1$, are updated by

$$\hat{g}_1^{(\nu)}(t; \hat{\boldsymbol{\beta}}_1, (y_i)_{i=1,\ldots,n}) = S^{(\nu)}(t, b_\nu; (\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_1, y_i)_{i=1,\ldots,n}). \tag{5}$$

2. (*Nonparametric estimation step for variance function*) Given $\hat{\boldsymbol{\beta}}_1$ and $\hat{g}_1(\cdot)$, an updated

nonparametric variance function estimate $\hat{\sigma}^2(\cdot)$ is obtained by

$$\hat{\sigma}^2(u) = S^{(0)}(u, b; (\hat{\mu}_i, \hat{\varepsilon}_i^2)_{i=1,\ldots,n}), \tag{6}$$

where $\hat{\mu}_i = \hat{g}_1(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_1)$ and $\hat{\varepsilon}_i^2 = (y_i - \hat{\mu}_i)^2$ are squared residuals which serve as the "raw" variance estimates and are based on the current model fit. Note that $\hat{\mu}_i = \sum_{j=1}^{M} \hat{g}_j(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_j)$ in the case $M > 1$.

3. (*Parametric estimation step*) Given $\hat{g}_1^{(0)}(\cdot)$, $\hat{g}_1^{(1)}(\cdot)$ and $\hat{\sigma}^2(\cdot)$, $\hat{\boldsymbol{\beta}}_1$ is updated by solving the following estimated estimating equation with respect to $\boldsymbol{\beta}_1$:

$$U(\boldsymbol{\beta}_1; \hat{g}_1^{(\nu)}(\cdot), \hat{\sigma}^2(\cdot)) = \sum_{i=1}^{n} \frac{y_i - \hat{g}_1^{(0)}(\eta_{1_i})}{\hat{\sigma}^2(\hat{g}_1^{(0)}(\eta_{1_i}))} \hat{g}_1^{(1)}(\eta_{1_i}) \boldsymbol{x}_i = 0, \tag{7}$$

where $\eta_{1_i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_1$ and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$. Denoting the root of (7) by $\tilde{\boldsymbol{\beta}}_1$, we then obtain the updated $\hat{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_1 / \|\tilde{\boldsymbol{\beta}}_1\|$.

The iteration continues until some convergence criterion is met. To simplify the notation, we denote the estimated vector of regression parameters and the estimated link and variance functions by

$$(\hat{\boldsymbol{\beta}}_1, \hat{g}_1(\cdot), \hat{\sigma}^2(\cdot)) = QLUE((\boldsymbol{x}_i, y_i)_{i=1,\ldots,n}). \tag{8}$$

3.2. Initial estimation step

Turning to the MUSE model, we sequentially apply the basic QLUE algorithm to each of the M indices ($M > 1$).

(I.1) For the first index $k = 1$, we fit the following approximate model with a single index:

$$y_i = g_1(\boldsymbol{x}_i^T \boldsymbol{\beta}_1) + \varepsilon_{1_i}, \tag{9}$$

by the QLUE algorithm, obtaining estimates (8).

(I.2) Set $k \leftarrow k+1$. If $k < M$, obtain $\tilde{\varepsilon}_{k_i} = y_i - \sum_{j=1}^{k-1} \hat{g}_j(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_j)$ as current pseudo-observations, and fit the model:

$$\tilde{\varepsilon}_{k_i} = g_k(\boldsymbol{x}_i^T \boldsymbol{\beta}_k) + \varepsilon_{k_i}, \tag{10}$$

again based on the QLUE algorithm, obtaining estimates $(\hat{\boldsymbol{\beta}}_k, \hat{g}_k(\cdot), \hat{\sigma}^2(\cdot)) = QLUE((\boldsymbol{x}_i, \tilde{\varepsilon}_{k_i})_{i=1,\ldots,n})$.

(I.3) Repeat step (I.2) until $k = M$.

We note that the resulting estimates might not be efficient because the fitted models at each step may not be close to the true model, thus allowing contamination of the subsequent parameter estimates. We therefore consider additional iteration steps using backfitting.

3.3. Backfitting algorithm

We observe that

$$E(y - \sum_{\substack{1 \le j \le M \\ j \ne k}} g_j(\boldsymbol{x}^T \boldsymbol{\beta}_j)) = g_k(\boldsymbol{x}^T \boldsymbol{\beta}_k),$$

suggests backfitting to control for biases as it is always based on residuals for which the assumed model holds. Define for $i = 1, \ldots, n$, and $j = 1, \ldots, M$,

$$z_{k_i} = y_i - \sum_{\substack{1 \le j \le M \\ j \ne k}} g_j(\eta_{j_i}), \tag{11}$$

where $\eta_{j_i} = \boldsymbol{x}_i^T \boldsymbol{\beta}_j$, so that $z_{k_i} = g_k(\boldsymbol{x}_i^T \boldsymbol{\beta}_k) + \varepsilon_i$ with $E\varepsilon_i = 0$. The data-based version is

$$\hat{z}_{k_i} = y_i - \sum_{\substack{1 \le j \le M \\ j \ne k}} \hat{g}_j(\hat{\eta}_{j_i}; \hat{\boldsymbol{\beta}}_j, (\hat{z}_{j_s})_{s=1,\ldots,n}), \tag{12}$$

where $\hat{\eta}_{j_i} = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_j$, and $\hat{\boldsymbol{\beta}}_j$ and $\hat{g}_j(\cdot)$ are obtained by the QLUE method, substituting "responses" $\hat{z}_{j_s}$, using current values from the previous iteration step. Denoting the ν-th derivative of the link function g_k ($\nu = 0$ or $\nu = 1$) by $g_k^{(\nu)}$, $k = 1, \ldots, M$, the backfitting estimation procedure then proceeds as follows.

(B.1) Set $k = 1$. Use the $\hat{z}_{k_i}$ in (12) as "observations" z_{k_i}, where $\hat{g}_j(\cdot)$, $j \neq k$, are obtained from the previous iteration steps. Then, fit model (11) by the QLUE algorithm to obtain the QLUE estimates $(\hat{\boldsymbol{\beta}}_k, \hat{g}_k(\cdot), \hat{\sigma}^2(\cdot)) = QLUE((\boldsymbol{x}_i, \hat{z}_{k_i})_{i=1,\ldots,n})$.

(B.2) Set $k \leftarrow k + 1$ and repeat step (B.1) until $k = M$.

(B.3) Repeat steps (B.1) and (B.2) until some convergence criterion is met. The convergence criterion we use in our implementation is

$$\sum_{j=1}^{M} \frac{\|\hat{\boldsymbol{\beta}}_j^{(\tau+1)} - \hat{\boldsymbol{\beta}}_j^{(\tau)}\|}{\|\hat{\boldsymbol{\beta}}_j^{(\tau)}\|} \Big/ \{M\ p\} \ < \ \epsilon \ ,$$

where the superscript (τ) indicates the τ-th backfitting loop and ϵ is a small threshold.

We note that the variance function is updated at each backfitting step for all indices based on the squared residuals as the raw variance estimates.

3.4. Bandwidth selection

When applying QLUE, which is the basic building block of the proposed method, the link and variance functions are estimated nonparametrically via smoothing techniques. Therefore, bandwidth selectors are required when applying the proposed methods. In previous work, Chiou & Müller (1998) demonstrated that bandwidth selectors based on the "nonparametric" quasi-deviance or the Pearson chi-square statistic work reasonably well in practice within the quasi-likelihood regression model with unknown link and variance functions and a single predictor. In this paper, we propose two alternative methods for bandwidth selection, which are based on extended quasi-likelihood (Nelder & Pregibon, 1987) respectively pseudo-likelihood (Ruppert & Carroll, 1988). Replacing link (mean) and variance functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ with their respective estimates $\hat{\mu}(\cdot)$ and $\hat{\sigma}^2(\cdot)$, we define a "nonparametric" extended quasi-likelihood as

$$Q^+(y; \hat{\mu}, \hat{\sigma}^2(\cdot)) = -\frac{1}{2} \sum_{i=1}^{n} \log\{2\pi\hat{\sigma}^2(y_i)\} - \frac{1}{2} D(y; \hat{\mu}, \hat{\sigma}^2(\cdot)), \tag{13}$$

where

$$D(y;\hat{\mu},\hat{\sigma}^2(\cdot)) = \sum_{i=1}^{n}\left\{-2\int_{y_i}^{\hat{\mu}_i}\frac{y_i - t}{\hat{\sigma}^2(t)}dt\right\} \tag{14}$$

is the "nonparametric" quasi-deviance. The bandwidths b_μ and b_σ, respectively, for link and variance functions, are then obtained as

$$(\hat{b}_\mu, \hat{b}_\sigma) \;= \underset{\{b_\mu, b_\sigma\}}{\arg\max}\; Q^+(y;\hat{\mu}_{b_\mu}, \hat{\sigma}^2_{b_\sigma}(\cdot)). \tag{15}$$

Pseudo-likelihood is motivated by a log-likelihood based on the normality assumption, and we adapt this concept by defining a "nonparametric" pseudo-likelihood

$$PL(y;\hat{\mu},\hat{\sigma}^2(\cdot)) = -\frac{1}{2}\sum_{i=1}^{n}\log\{2\pi\hat{\sigma}^2(\mu_i)\} - \frac{1}{2}P(y;\hat{\mu},\hat{\sigma}^2(\cdot)), \tag{16}$$

where

$$P(y;\hat{\mu},\hat{\sigma}^2(\cdot)) = \sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}^2(\hat{\mu}_i)} \tag{17}$$

is the "nonparametric" Pearson statistic. The selected bandwidths are then

$$(\hat{b}_\mu, \hat{b}_\sigma) \;= \underset{\{b_\mu, b_\sigma\}}{\arg\max}\; PL(y;\hat{\mu}_{b_\mu}, \hat{\sigma}^2_{b_\sigma}(\cdot)). \tag{18}$$

When the variance function is unknown and needs to be estimated from the data, the quasi-deviance D (14) and Pearson's statistics P (17), that respectively appear in the second term of Q^+ (15) and PL (18), both increase as variances decrease. The influence of the bandwidth on the variance function estimate thus negotiates a compromise between the first and the second terms in both Q^+ (15) and PL (18). A more general interpretation is that these criteria amount to a penalized deviation.

In the estimation procedure for the MUSE model components, the link and the variance functions are estimated alternatingly. Thus both the above two criteria for bandwidth selection can be used for estimating the link and the variance functions, and these two criteria can be used alternatively; for instance, the bandwidth selector for link function

estimation could be based on maximizing extended quasi-likelihood, treating variance function estimates as fixed, while the bandwidth selector for variance function could be based on maximizing pseudo-likelihood, treating link function estimates as fixed.

Since both the "nonparametric" extended quasi-likelihood and pseudo-likelihood are highly nonlinear in the bandwidths, a heuristic approach by grid search is used for practical implementation. Alternative bandwidth selectors include AIC- or BIC-like criteria. Previous finite sample studies of QLUE reveal that these criteria lead to similar results.

3.5. Choice of number of indices M

For the choice of M, the number of indices within the MUSE model, we propose an ad hoc sequential data-driven approach based on a penalized sum of squared errors (PSE). Given a number of indices $m \geq 1$, PSE is defined as

$$PSE(m) = SSE(m) + m\ p\ MSE(m), \tag{19}$$

where $\mathrm{SSE}(m) = \sum_{i=1}^{n} \left(y_i - \sum_{k=1}^{m} \hat{g}_k(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_k) \right)^2$ and $\mathrm{MSE}(m) = \mathrm{SSE}(m)/n$. The second term penalizes against larger numbers of indices in the model. In analogy to Mallow's Cp statistic, the penalty could also be doubled, leading to the alternative criterion

$$PSE^*(m) = SSE(m) + 2\ m\ p\ MSE(m). \tag{20}$$

The chosen number of indices M is then the minimizer of $PSE(m)$ or $PSE^*(m)$.

An obvious alternative criterion is cross-validation or leave-one-out squared prediction error. However, these criteria are considerably more computing intensive than those discussed above.

4. Constrained estimation in a MUSE model with orthogonal indices

In the proposed algorithm, the regression parameter vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M$ defining the indices are estimated iteratively with backfitting. The estimates of the regression parameter

vectors satisfy $\|\hat{\boldsymbol{\beta}}_k\| = 1$, according to (M6). In some applications, additional restrictions such as orthogonality of the parameter vector estimates may be appropriate, either from the nature of the problem at hand or for enhancing the stability of the estimates, especially in the case of highly correlated indices. Formally, an additional orthogonality constraint is

(M7) The regression parameter vectors $\boldsymbol{\beta}_k$ in (1) are orthogonal to one another, i.e.,

$$\boldsymbol{\beta}_j^T \boldsymbol{\beta}_\ell = 0 \ , \ \ 1 \leq j \neq \ell \leq M.$$

To facilitate the notation, let F_k, $k = 1, \ldots, M$, be fixed $p \times (M-1)$ matrices whose columns represent the coefficients of a linear constraint on the parameter vectors $\boldsymbol{\beta}_k$,

$$F_k^T \boldsymbol{\beta}_k = 0, \ k = 1, \ldots, M. \tag{21}$$

The solution of the estimating equations $U(\boldsymbol{\beta}_k; \hat{g}_k^{(\nu)}, \hat{\sigma}^2(\cdot)) = 0$ (7) under constraints (21) for $\boldsymbol{\beta}_k$ and also the asymptotic variance of the constrained estimates, differ from the unconstrained case. In the cases of orthogonality constraints, the column vectors of the matrices F_k are the vectors $\boldsymbol{\beta}_j$ for $j \neq k$. Define $p \times p$ matrices

$$\boldsymbol{i}_{\beta_k} = D_k^T V^{-1} D_k, \quad k = 1, \ldots, M, \tag{22}$$

as in (M5). We note that $\boldsymbol{i}_{\beta_k}$ is the expected information matrix for $\boldsymbol{\beta}_k$ in the estimating equation $U(\boldsymbol{\beta}_k; \hat{g}_k(\cdot), \hat{\sigma}^2(\cdot)) = 0$ (7). Under linear constraints (21), consider projection matrices

$$P_k = F_k (F_k^T \boldsymbol{i}_{\beta_k}{}^{-1} F_k)^{-1} F_k^T \boldsymbol{i}_{\beta_k}{}^{-1}, \ k = 1, \ldots, M. \tag{23}$$

For $M = 1$, under correct link and variance function specification as in the conventional quasi-likelihood model satisfying the linear constraints (21), the projected estimator is obtained as $\tilde{\boldsymbol{\beta}}_k^* = (I_p - P_k)^T \tilde{\boldsymbol{\beta}}_k$, where $\tilde{\boldsymbol{\beta}}_k$ is the conventional quasi-likelihood estimator and I_p is the $p \times p$ identity matrix. It can be easily verified that $\tilde{\boldsymbol{\beta}}_k^*$ indeed satisfies the linear

constraint $F_k^T \tilde{\boldsymbol{\beta}}_k^* = 0$. Note that Heyde & Morton (1993) referred to this method as "projection of free parameters" and showed that using the projection matrix (23) leads to the asymptotic variance $cov(\tilde{\boldsymbol{\beta}}_k^*) \approx (I_p - P_k)^T \boldsymbol{i}_{\beta_k}^{-1} (I_p - P_k)$.

However, in the MUSE model the situation is more complex, as both link and variance functions are unknown and need to be estimated nonparametrically. The unknowns $g_k(\cdot)$ and $\sigma^2(\cdot)$ in $\boldsymbol{i}_{\beta_k}$ (22) are replaced by the estimates $\hat{g}_k(\cdot)$ and $\hat{\sigma}^2(\cdot)$ (5) and (6), leading to

$$\hat{\boldsymbol{i}}_{\beta_k} = \hat{D}_k^T \hat{V}^{-1} \hat{D}_k, \tag{24}$$

where $\hat{V}$ is a diagonal $n \times n$ matrix with the ith diagonal element $\hat{\sigma}^2(\hat{\mu}_i)$ and $\hat{D}_k$ is a $n \times p$ matrix with the (i,r)th element $(\hat{D}_k)_{ir} = \hat{g}_k^{(1)}(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_k) x_{ir}$, where the $\hat{\boldsymbol{\beta}}_k$'s are the solutions of the estimating equations $U(\boldsymbol{\beta}_k; \hat{g}_k(\cdot), \hat{\sigma}^2(\cdot)) = 0$ (7). Consequently, the projection matrices are

$$P_k^* = F_k (F_k^T \hat{\boldsymbol{i}}_{\beta_k}^{-1} F_k)^{-1} F_k^T \hat{\boldsymbol{i}}_{\beta_k}^{-1}, \tag{25}$$

and the projected estimates $\hat{\boldsymbol{\beta}}_k^*$ are obtained from $\hat{\boldsymbol{\beta}}_k$ by

$$\hat{\boldsymbol{\beta}}_k^* = (I_p - P_k^*)^T \hat{\boldsymbol{\beta}}_k, \tag{26}$$

for $k = 1, \ldots, M$.

The covariance matrix of the estimates $\hat{\boldsymbol{\beta}}_k^*$ is obtained by applying the delta method, which will be discussed in section 6. In the MUSE model, the projection steps are embedded in the QLUE procedure while updating $\hat{\boldsymbol{\beta}}_k$, for $1 < k \leq M$, so that link and variance functions as well as the (projected) regression coefficients are updated at each QLUE iteration.

5. Data applications

The proposed MUSE model and estimation procedures are illustrated with two real data sets: one from a food folate experiment in a rat growth bioassay and the other from a medfly reproduction experiment with various feeding schemes.

5.1. Folate bioassay

The folate data were previously analyzed by Müller *et al.* (1996). The motivation behind the experiment of a folate depletion-repletion rat growth bioassay was to compare how several sources of dietary folate affect the growth rates of folate-depleted rats. In Müller *et al.* (1996), a linear statistical interaction model for predicting the growth-promoting effect of several sources of dietary folate was developed. It was concluded that food folates generally are not exchangeable and do interact adversely. In the MUSE analysis, we use the three different folate sources as predictors: Folic Acid (FA: pure compound folic acid), Bean Folate (BF: folate from cooked pinto beans) and Liver Folate (LF: folate from cooked fried beef liver); response is weight growth (g/day).

The iterative backfitting procedure based on the QLUE approach was used with the orthogonality constraint (M7) in place. We first fitted models with various numbers of M and chose the best number of link functions according to the two penalized SSE criteria PSE (19) and PSE^* (20). These two criteria results in different choices (Table 1): $M = 2$ chosen by PSE (19) and $M = 1$ by PSE^* (20). The smoothing parameters for estimating link and variance functions are selected via extended quasi-likelihood (15). The fitted model for $M = 2$ is presented in Table 2 and Fig. 1. We note that in the backfitting algorithm, the value of the tolerance ϵ in step (B.3) was $\epsilon = 0.001$ and convergence occurred after 3 iterations.

The left panel of Fig. 1 reveals that the effect of the first index, which is essentially an average over all three folate sources (see Table 2), reflects a saturation effect. Such effects are common in nutritional studies. The inclusion of nonparametric link function estimates is essential here for the detection of this effect. The right panel of Fig. 1 demonstrates that the second index, which is essentially a contrast between bean and liver folate levels (see Table 2), has as smaller modulating effect, enhancing growth for smaller index levels that

correspond to a higher liver folate/bean folate ratio.

Table 2 provides the fitted regression coefficients with their standard errors, the goodness-of-fit statistics and the values of smoothing parameters selected for each link function. Although bean folate (BF) and liver folate (LF) have nearly the same effect on growth rate in the first index, they figure in the second index with opposite signs. The second index therefore can be viewed as a contrast between BF on one and FA, LF on the other side. An implication is that bean folate (BF) and liver folate (LF) interact adversely, thus confirming the earlier conclusion that food folates generally are not exchangeable.

Insert Table 1 about here!

Insert Table 2 about here!

Insert Figure 1 about here!

The nonparametric goodness-of-fit statistics (G.O.F.) for model fits in the iterative back-fitting procedure based on residuals include "nonparametric" quasi-deviance D as in (14) and the Pearson statistic P as in (17). Details about these statistics can be found in Chiou & Müller (1998). Their asymptotic expected values are the same, and in a good model fit these values should be similar. This indeed seems to be the case here.

5.2. Medfly reproduction in response to diet

Our second data example is from a medfly (Ceratitis capitata) reproduction experiment in which a variety of protein-sugar feeding schemes were applied to 645 medflies (three outliers were excluded from the analysis). The medfly data were previously analyzed by Carey *et al.* (2002). The purpose of this experiment is to examine the influence of various dietary schemes on the reproduction of medflies as measured in terms of total number of eggs produced, which serves as a proxy for evolutionary fitness. A factor that needs to be

taken into account as well is the lifetime of a medfly, which will obviously have an effect on the number of eggs produced, as longer living flies are enabled to produce more eggs.

The various feeding schemes of the experiment are listed in Table 3. The relevant variables are listed in Table 4. Treatment 1 is viewed as the baseline and we define the other treatments by a set of indicator variables.

Insert Table 3 about here!

Insert Table 4 about here!

In this example, the bandwidths chosen for estimating link and variance functions are based on the method of minimizing pseudo-likelihood (18). We chose M by fitting models with various numbers of indices m and calculating the penalized SSE, PSE (19) and PSE^* (20). The results are in Table 5, indicating the optimal choice is $M = 4$ for all criteria. The resulting fitted model with $M = 4$ is presented in Table 6 and Fig. 2. We note that the iteration over the backfitting steps terminates after four iterations. As in the first example, the orthogonality constrained version of the algorithm was implemented.

Insert Table 5 about here!

Insert Table 6 about here!

The estimated regression coefficients defining the indices as shown in Table 6 reveal that the first index increases with diets that contain protein, and the second index increases is closely tied to the full protein diet. The third index becomes negative for more sporadic protein treatments and increases for the regular protein treatments. Finally, the fourth index in a zig-zag pattern becomes larger for the 1:3 and 1:10 protein diets, and declines for the 1:5 and 1:20 protein diets.

The effects of increasing each of the four indices on medfly reproduction can be seen from the link functions that are shown in Fig. 2. We find that increasing the value of the first index leads to a steady increase in reproduction up to a certain level of the index, while increasing the second index leads to a steadily increasing and then decreasing response. Increasing the third index leads to an initially flat and then strongly increasing response. The fourth index behaves differently in that an increase in that index is associated with a decline in egg output. This index represents a decline in reproduction for long lived flies who are fed small amounts of protein, hinting at a trade-off between reproduction and longevity.

Insert Figure 2 about here!

Insert Figure 3 about here!

As another illustration of the fitted model, Fig. 3 demonstrates the treatment effects on the relationship between reproduction and lifetime. The solid curves represent the predicted relationship, overlaid with the corresponding raw data for the specific treatment, and the dashed curves represent the relationship under baseline treatment (sugar only). A saturation effect in terms of reproduction is evident for long-lived flies in all treatment groups and occurs after about 60 days. After saturation, increased longevity does not lead to increased reproduction. This example illustrates the multiple index model for a case where categorical predictor variables play a major role.

6. Asymptotic properties

We aim to show that the link function estimates of $g_k(\cdot)$, $k = 1, \ldots, M$, and the variance function estimate of $\sigma^2(\cdot)$ are consistent, and that the resulting estimates of the regression parameters as well as the "projected" regression parameter estimates are asymptotically normally distributed, as in conventional quasi-likelihood regression models when

link and variance functions are correctly specified. Let $\mathcal{B}_n^j = \{\bar{\boldsymbol{\beta}}_j : \sqrt{n}\ \|\bar{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\| = B \text{ for some } B \text{ with } 0 < B < \infty\}$, for $j = 1, \ldots, M$, where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^p$. Define

$$\tilde{z}_{k_i} = y_i - \sum_{\substack{1 \le j \le M \\ j \ne k}} \hat{g}_j(\tilde{\eta}_{j_i}; \bar{\boldsymbol{\beta}}_j, (\tilde{z}_{j_s})_{s=1,\ldots,n}), \tag{27}$$

for $k = 1, \ldots, M$, and $i = 1, \ldots, n$, where $\tilde{\eta}_{j_i} = \boldsymbol{x}_i^T \bar{\boldsymbol{\beta}}_j$, $\bar{\boldsymbol{\beta}}_j \in \mathcal{B}_n^j$, $\hat{g}_j(\cdot)$ are defined in (5) and $\tilde{z}_{j_s}$ are iteratively defined in (27). We further let I_μ and I_{η_k}, $k = 1, \ldots, M$, be the compact intervals which represent the interiors of the supports of the "design density" of the means and the kth linear predictors, respectively. Theorem 1 provides the basic consistency result for the nonparametric link function estimates as well as for their first derivatives which are used in the auxiliary step of estimating $\boldsymbol{\beta}_k$. The consistency result for the nonparametrically estimated variance function is presented in theorem 2. The proofs of the following theorems are compiled in the appendix.

Theorem 1

Suppose that $\bar{\boldsymbol{\beta}}_j \in \mathcal{B}_n^j$, $j = 1, \ldots, M$, and for an arbitrary index k, $1 \le k \le M$, that $\max_{\substack{1 \le j \le M \\ j \ne k}} \max_{1 \le i \le n} |\tilde{z}_{j_i} - z_{j_i}| = O_p(\delta_n)$, and $\delta_n \to 0$ as $n \to \infty$, where z_{j_i} and $\tilde{z}_{j_i}$ are defined in (11) and (27), respectively. If there exists a sequence $\bar{\boldsymbol{\beta}}_k = \bar{\boldsymbol{\beta}}_{kn}$ such that $\sqrt{n}\ \|\bar{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = O_p(1)$, then under (M1)-(M6) and the regularity conditions for smoothing, for $\nu = 0, 1$,

$$\sup_{t \in I_{\eta_k}} |\hat{g}_k^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k, (\tilde{z}_{k_i})_{i=1,\ldots,n}) - g_k^{(\nu)}(t)| = O_p(\gamma_{n\nu} + \delta_n), \tag{28}$$

where $\gamma_{n\nu} = 1/\sqrt{n} b_{k\nu}^{1+\nu} + \{\log n / n b_{k\nu}^{2\nu+1}\}^{1/2} + b_{k\nu}^2$, and $b_{k\nu}$ is the bandwidth used for the nonparametric estimation of $\hat{g}_k^{(\nu)}$, as defined in (5).

Theorem 2

If there exist sequences $\bar{\boldsymbol{\beta}}_k = \bar{\boldsymbol{\beta}}_{kn}$ *such that* $\sqrt{n}\ \|\bar{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = O_p(1)$, $k = 1, \ldots, M$, *then under (M1)-(M6) and the regularity conditions for smoothing,*

$$\sup_{u \in I_\mu} |\hat{\sigma}^2(u) - \sigma^2(u)| = O_p(\boldsymbol{\alpha}_n + (\gamma_{n0} + \delta_n) b^{-1}), \tag{29}$$

where $\boldsymbol{\alpha}_n = \{\log n/nb\}^{1/2} + b^2$, γ_{n0} *and* δ_n *are defined in theorem 1, and* b *is the bandwidth used for nonparametric estimation of* $\hat{\sigma}^2(\cdot)$, *as defined in (6).*

The main results demonstrating the asymptotic efficient adaptive estimation for the regression parameter vectors and also for the estimates under orthogonal constraint (M7) are presented in the following theorem. We note that because of the assumption $\|\boldsymbol{\beta}_k\| = 1$ in (M6) and the normalization step in the estimated vector of regression parameters, the constrained p-vectors $\boldsymbol{\beta}_k$ are actually in $\mathbb{R}^{p-1}$, and therefore the covariance matrix of the estimated parameter vector $\hat{\boldsymbol{\beta}}_k$ is of rank $(p-1)$. The adjustment of the estimated covariance matrix for $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\beta}}_k^*$ is therefore included in the following theorem.

To facilitate the notations, we define functions $f_i : \mathbb{R}^p \to \mathbb{R}$, $1 \le i \le p-1$, such that $f_i(\mathbf{u}) = u_i/\|\mathbf{u}\|$, where $\mathbf{u} = (u_1, \ldots, u_p)^T$, and a function $f : \mathbb{R}^p \to \mathbb{R}^{p-1}$ such that

$$\begin{aligned} f(\mathbf{u}) &= (f_1(\mathbf{u}), \ldots, f_{p-1}(\mathbf{u}))^T \\ &= (u_1/\|\mathbf{u}\|, \ldots, u_{p-1}/\|\mathbf{u}\|)^T. \end{aligned}$$

Furthermore, we define a $p \times (p-1)$ matrix $(Df)(\mathbf{u})$ as

$$(Df)(\mathbf{u}) = \left(\frac{\partial f_i(\mathbf{u})}{\partial u_j} \right)_{1 \le i \le p-1, 1 \le j \le p}. \tag{30}$$

Based on the notations defined above, the asymptotic distribution of $f(\tilde{\boldsymbol{\beta}}_k) = (\hat{\boldsymbol{\beta}}_{k_1}, \ldots, \hat{\boldsymbol{\beta}}_{k_{p-1}})^T$ and $f(\tilde{\boldsymbol{\beta}}_k^*) = (\hat{\boldsymbol{\beta}}_{k_1}^*, \ldots, \hat{\boldsymbol{\beta}}_{k_{p-1}}^*)^T$ is of interest.

Theorem 3

In the MUSE model, under (M1)-(M6) and the regularity conditions on smoothing, if the link function and its first derivative are estimated by $\hat{g}_k^{(\nu)}$ *in (5), and the variance function is estimated by* $\hat{\sigma}^2(\cdot)$ *in (6) which is truncated below so that it is bounded away from 0, then for a given arbitrarily small* $\zeta > 0$*, the following holds on an event with probability* $1 - \zeta$*: There exists a QLUE* $\tilde{\boldsymbol{\beta}}_k$ *of (7), and* $f(\tilde{\boldsymbol{\beta}}_k)$ *is asymptotically normally distributed such that, as* $n \to \infty$*,*

$$\sqrt{n}\,(f(\tilde{\boldsymbol{\beta}}_k) - f(\boldsymbol{\beta}_k)) \xrightarrow{\mathcal{D}} N_p(\mathbf{0},\ [(Df)(\boldsymbol{\beta}_k)]\,\Sigma_k^{-1}\,[(Df)(\boldsymbol{\beta}_k)]^T),\ k = 1, \ldots, M, \tag{31}$$

where $\Sigma_k = \lim_{n\to\infty} \frac{1}{n} D_k^T V^{-1} D_k$ *as defined in (M6) and* $(Df)(\boldsymbol{\beta}_k)$ *as in (30). In addition, under (M7) the projected QLUE estimates* $\tilde{\boldsymbol{\beta}}_k^*$ *(26) are asymptotically normally distributed such that, as* $n \to \infty$*,*

$$\sqrt{n}\,(f(\tilde{\boldsymbol{\beta}}_k^*) - f(\boldsymbol{\beta}_k^*)) \xrightarrow{\mathcal{D}} N_p(\mathbf{0},\ [(Df)(\boldsymbol{\beta}_k^*)]\,\Sigma_k^{*-1}\,[(Df)(\boldsymbol{\beta}_k^*)]^T), \tag{32}$$

where $\boldsymbol{\beta}_k^* = (I_p - P_k)\boldsymbol{\beta}_k$ *and* $\Sigma_k^* = \lim_{n\to\infty} \frac{1}{n}(I_p - P_k)^T (D_k^T V^{-1} D_k)^{-1}(I_p - P_k)$*.*

Theorem 3 implies the following result which is of practical importance.

Corollary 1

Under the assumptions in theorem 3, where $(Df)(\tilde{\boldsymbol{\beta}}_k)$ *and* $(Df)(\tilde{\boldsymbol{\beta}}_k^*)$ *are defined according to (30), let*

$$\hat{\Sigma}_k^{-1} = n\,(\hat{D}_k^T \hat{V}^{-1} \hat{D}_k) \quad \text{and} \quad \hat{\Sigma^*}_k^{-1} = (I_p - P_k^*)^T \hat{\Sigma}_k^{-1} (I_p - P_k^*),$$

where P_k^* *is as in (24), (25),* $\hat{V} = diag\left(\{\hat{\sigma}^2(\hat{\mu}_i)\}\right)_{1\le i\le n}$ *with* $\hat{\mu}_i = \sum_{k=1}^{M} \hat{g}_k(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_k)$*, and* $\hat{D}_k = (\hat{D}_k)_{ir\ \substack{1\le i\le n \\ 1\le r\le p}}$ *with* $(\hat{D}_k)_{ir} = \hat{g}_k^{(1)}(\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_k)\, x_{ir}$*. Then,*

$$(Df)(\tilde{\boldsymbol{\beta}}_k) \xrightarrow{p} (Df)(\boldsymbol{\beta}_k), \quad \hat{\Sigma}_k \xrightarrow{p} \Sigma_k, \quad \text{and}$$

$$f(\tilde{\boldsymbol{\beta}}_k) \dot{\sim} N_p\left(f(\boldsymbol{\beta}_k), \frac{1}{n}[(Df)(\tilde{\boldsymbol{\beta}}_k)]\hat{\Sigma}_k{}^{-1}[(Df)(\tilde{\boldsymbol{\beta}}_k)]^T\right), \tag{33}$$

for large n. Similarly,

$$(Df)(\tilde{\boldsymbol{\beta}}_k^*) \xrightarrow{p} (Df)(\boldsymbol{\beta}_k^*), \quad \hat{\Sigma_k^*} \xrightarrow{p} \Sigma_k^* \quad \textit{and}$$

$$f(\tilde{\boldsymbol{\beta}}_k^*) \dot{\sim} N_p\left(f(\boldsymbol{\beta}_k^*), \frac{1}{n}[(Df)(\tilde{\boldsymbol{\beta}}_k^*)]\hat{\Sigma_k^*}{}^{-1}[(Df)(\tilde{\boldsymbol{\beta}}_k^*)]^T\right), \tag{34}$$

for large n.

To have a closer look at the estimated asymptotic covariance matrix in (33) above, define the $(p-1)\times(p-1)$ matrix

$$\mathbf{V}_k' = [(Df)(\tilde{\boldsymbol{\beta}}_k)]\ \hat{\Sigma}_k^{-1}\ [(Df)(\tilde{\boldsymbol{\beta}}_k)]^T$$

with elements v_{rs}' for $1 \le r, s \le (p-1)$. The elements can be expressed explicitly as

$$v_{rs}' = \sum_{1\le i,j\le p} \frac{1}{\|\tilde{\boldsymbol{\beta}}_k\|^2}\{1_{\{i=r\}} - \tilde{\boldsymbol{\beta}}_{k_r}\tilde{\boldsymbol{\beta}}_{k_i}\}\{1_{\{j=s\}} - \tilde{\boldsymbol{\beta}}_{k_s}\tilde{\boldsymbol{\beta}}_{k_j}\}\hat{\sigma}_{ij}, \tag{35}$$

where $\hat{\sigma}_{ij}$ is the (i,j)th element of $\hat{\Sigma}^{-1}$. In the proposed estimation scheme, the regression parameter estimates are normalized to one at each iteration of the QLUE method. The "adjusted" covariance estimates of the estimated vector of regression coefficients can then be calculated easily. We note that the expression in (35) is relevant for (33). Similarly, an explicit expression for (34) can be obtained by replacing $\hat{\sigma}_{ij}$ in (35) with $\hat{\sigma}_{ij}^*$, where $\hat{\sigma}_{ij}^*$ is the (i,j)th element of $\hat{\Sigma}^{*-1}$.

7. Concluding remarks

We have introduced a highly flexible semi-parametric regression model which includes multiple indices and provides a general modeling framework for regression data. Discrete indicator predictor variables can be easily included and the extension to binomial and other

types of response variables can be implemented as well. The proposed model is an extension of quasi-likelihood GLM approaches. It provides an extension from commonly used single index models, to the case of multiple indices where link and variance functions are unknown but smooth. The proposed model is especially useful when the assumption of a single linear predictor is not appropriate or when complex interactions exist. The proposed model can also be viewed as a nonparametric extension of principal components regression, where the dependent variable is used to orthogonalize the indices.

The classical backfitting method of Hastie & Tibshirani (1990), along with its modifications, has been used as the main tool for fitting a variety of additive models. It is intuitive, easy to implement and works well in simulation studies and applications with real data. However, the classical backfitting method may not be design-adaptive, especially when the covariates are highly correlated, as was pointed out by Mammen, Linton & Nielsen (1999). These authors proposed a new backfitting-type estimator for additive nonparametric regression, which they proved to be design-adaptive, and it would be of interest in future work to explore the implementation of this new algorithm for the MUSE model.

The proposed MUSE model serves as a tool for dimension reduction; in most applications only few indices will be needed. In addition to the PSE (19) and PSE^* (20) criteria for selecting the number of indices, other selection methods can be based on visual inspection of link functions, minimizing cross-validation prediction errors, or statistical tests based on appropriately constructed confidence bands. Moreover, for each index one may apply variable selection techniques, which also may be based on the proposed penalized square error criteria PSE, PSE^*. This induces additional iterations by dropping predictors in individual indices, and then refitting the indices under new constraints for a reduced number of predictors. Here the projection method of constrained estimation comes in very handy to enforce both dimension reduction and shrinkage.

Another note regarding fitting probabilities for data with binomial responses is that in such situations one needs to ensure that the fitted values fall between 0 and 1, especially for fitting a model of additive type as the MUSE model. There are several possibilities to enforce this. One way is to include a known overall link function and then to fit the additive indices within the linear predictor for this overall link function. This is the approach which has been implemented in GAM (Generalized additive models) and it is straightforward to include it with MUSE. An alternative and possibly better way is to fit the model to log odds ratios instead of relative frequency. This is often of interest in itself and has the advantage of not imposing a range constraint for the fitted odds when dealing with binomial responses. Efficiency is guaranteed by our results as the data themselves determine the variance function. In a last step, it is then easy to transform the fitted odds to the corresponding probabilities and the resulting fitted probabilities will fall between 0 and 1.

We have demonstrated in the two example data sets that the MUSE model is suitable for the modeling of data with complex interactions and can lead to interesting insights.

Acknowledgments

We wish to thank two referees, an Associate Editor and the Editor for very helpful remarks that led to many improvements in the paper. This research was supported by NHRI grant BS-091-PP07, NSC grant 89-2118-M-194-004, NIH grant P01-08761 and NSF grant DMS-02-04869.

References

Carey, J.R., Liedo, P., Harshman, L., Zhang, Y., Müller, H.G., Partridge, L. & Wang, J.L. (2002). Life history response of Mediterranean fruit flies to dietary restriction. *Aging Cell* **1**, 140-148.

Carroll, R.J., Fan, J. Gijbels, I. & Wand, M.P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Carroll, R.J. & Ruppert, D. (1988). *Transformation and weighting in regression.* Chapman & Hall, New York.

Chiou, J.-M. & Müller, H.-G. (1998). Quasi-likelihood regression with unknown link and variance functions. *J. Amer. Statist. Assoc.* **93**, 1376-1387.

Friedman, J.H. & Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 817-823.

Hastie, T. & Tibshirani, R. (1986). Generalized additive models (with discussion). *Statist. Sci.* **1**, 297-318.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models.* Chapman & Hall, London.

Heyde, C.C. & Morton, R. (1993). On constrained quasi-likelihood estimation. *Biometrika* **80**, 755-761.

Li, K.-C. (1991). Sliced inverse regression for Dimension Reduction. *J. Amer. Statist. Assoc.* **86**, 316-327.

Lingjærde, O. C. & Liestøl, K. (1998). Generalized projection pursuit regression. *SIAM J. Sci. Comput.* **20**, 844-857.

Mammen, E., Linton, O. & Nielsen, J. (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443-1490.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59-67.

Müller, H.-G., Facer, M. R., Bills, N.D. & Clifford, A.J. (1996). Statistical interaction model for exchangeability of food folates in a rat growth bioassay. *J. Nutrition* **126**, 2585-2592.

Nelder, J.A. & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* **74**, 221-232.

Roosen, C.B. & Hastie, T.J. (1994). Automatic smoothing spline projection pursuit. *Journal of Computational and Graphical Statistics* **3**, 235-248.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and Gauss-Newton method. *Biometrika* **61**, 439-447.

Zhang, B. (1999). A chi-squared goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **86**, 531-539.

Zhang, B. (2000). A goodness-of-fit test for multiplicative intercept risk models based on case-control data. *Statistica Sinica* **10**, 839-865.

Zhang, B. (2001). An information matrix test for logistic regression models based on case-control data. *Biometrika* **88**, 921-932.

Jeng-Min Chiou, Division of Biostatistics and Bioinformatics, National Health Research Institutes, 128 Yen-Chiu-Yuan Rd. Sec. 2, Taipei 115, Taiwan, R.O.C.

E-mail: jmchiou@nhri.org.tw

Appendix A. Proof of identifiability

First assume that $M = 1$ and that there exists a link function $\tilde{g}_1$ and a vector $\tilde{\boldsymbol{\beta}}_1 \in \mathbb{R}^p$, $\tilde{\boldsymbol{\beta}}_1 \neq \boldsymbol{\beta}_1$, such that $g_1(\boldsymbol{x}^T\boldsymbol{\beta}_1) = \tilde{g}_1(\boldsymbol{x}^T\tilde{\boldsymbol{\beta}}_1)$, for all $\boldsymbol{x}$. There obviously exist $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ such that $\boldsymbol{x}_1^T\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{x}_2^T\tilde{\boldsymbol{\beta}}_1$ but $\boldsymbol{x}_1^T\boldsymbol{\beta}_1 \neq \boldsymbol{x}_2^T\boldsymbol{\beta}_1$. But the monotonicity of g_1, $\tilde{g}_1$ (M1) implies then $\tilde{g}_1(\boldsymbol{x}_1^T\tilde{\boldsymbol{\beta}}_1) = \tilde{g}_1(\boldsymbol{x}_2^T\tilde{\boldsymbol{\beta}}_1)$ while $g_1(\boldsymbol{x}_1^T\boldsymbol{\beta}_1) \neq g_1(\boldsymbol{x}_2^T\boldsymbol{\beta}_1)$, a contradiction. Since the design is asymptotically dense according to (M3) and (M4), the true parameter vector $\boldsymbol{\beta}$ is identifiable. Since $\|\boldsymbol{\beta}\| = 1$, the true link function is also identifiable. For $M \geq 2$, we assume $\sum_{i=1}^{M} g_i(\boldsymbol{x}^T\boldsymbol{\beta}_i) = \sum_{i=1}^{M} \tilde{g}_i(\boldsymbol{x}^T\tilde{\boldsymbol{\beta}}_i)$ for an alternative set of parameters $\tilde{\boldsymbol{\beta}}_i$ and link functions $\tilde{g}_i$, which satisfy all conditions. Then we use (M1) to conclude as above that $g_i(\boldsymbol{x}_1^T\boldsymbol{\beta}_i) = g_i(\boldsymbol{x}_2^T\boldsymbol{\beta}_i)$, $1 \leq i \leq M$, for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in L_c$, where c is adapted to $\tilde{\boldsymbol{\beta}}_1, \ldots, \tilde{\boldsymbol{\beta}}_M$, and that $\tilde{g}_i(\boldsymbol{x}_1^T\tilde{\boldsymbol{\beta}}_i) = \tilde{g}_i(\boldsymbol{x}_2^T\tilde{\boldsymbol{\beta}}_i)$, $1 \leq i \leq M$, $i \neq i_0$, and $\tilde{g}_{i_0}(\boldsymbol{x}_1^T\tilde{\boldsymbol{\beta}}_{i_0}) \neq \tilde{g}_{i_0}(\boldsymbol{x}_2^T\tilde{\boldsymbol{\beta}}_{i_0})$, whence we again have a contradiction.

Appendix B. Proof of theorems

The link function estimates $\hat{g}_k^{(\nu)}(\cdot)$ can be expressed as

$$\hat{g}_k^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k, (\tilde{z}_{k_i})_{i=1,\ldots,n}) = \sum_{i=1}^{n} G_{k_i}^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k)\ \tilde{z}_{k_i} \tag{36}$$

where $G_{k_i}^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k)$ are the weight functions corresponding to the locally weighted least squares smoother, and $\bar{\boldsymbol{\beta}}_k$ and $\tilde{z}_{k_i}$ satisfy the conditions in theorem 1. Define the variables: $z_{k_i}^{(1)} = y_i - \sum_{1\le\ell\le M, \ell\neq k} \hat{g}_\ell(\eta_{\ell i}; \bar{\boldsymbol{\beta}}_\ell, (\tilde{z}_{\ell s})_{s=1,\ldots,n})$, $z_{k_i}^{(2)} = y_i - \sum_{1\le\ell\le M, \ell\neq k} \hat{g}_\ell(\eta_{\ell i}; \bar{\boldsymbol{\beta}}_\ell, (z_{\ell s})_{s=1,\ldots,n})$, with z_{k_i} and $\tilde{z}_{k_i}$ as in (11) and (27).

Lemma 1

Given the weight function $G_{k_i}^{(\nu)}$ in (36) and $t \in I_{\eta_k}$ corresponding to the support of the kernel function in $G_{k_i}^{(\nu)}$, if a random variable $\tilde{t}$ satisfies $|\tilde{t} - t| = O_p(\frac{1}{\sqrt{n}})$, then under (M4) and the regularity conditions on smoothing, $\sup_{t\in I_{\eta_k}} \max_{1\le i\le n} |G_{k_i}^{(\nu)}(\tilde{t}; \boldsymbol{\beta}_k) - G_{k_i}^{(\nu)}(t; \boldsymbol{\beta}_k)| = O_p(\frac{1}{n\sqrt{n} b_\nu^{2+\nu}})$.

Lemma 2

Under (M4) and the regularity conditions on smoothing, $\max_{1\le i\le n} |\tilde{z}_{k_i} - z_{k_i}^{(1)}| = O_p(\frac{1}{\sqrt{n} b_0})$, $\max_{1\le i\le n} |z_{k_i}^{(1)} - z_{k_i}^{(2)}| = O_p(\delta_n)$, and $\max_{1\le i\le n} |z_{k_i}^{(2)} - z_{k_i}| = O_p(\gamma_{n0})$, where δ_n and γ_{n0} as in theorem 1 and $\hat{g}_k$ as in (36) for some M, $1 \le k \le M$.

Proof of theorem 1. Observe

$$\sup_{t\in I_{\eta_k}} |\hat{g}_k^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k, (\tilde{z}_{k_i})_{i=1,\ldots,n}) - g_k^{(\nu)}(t)| \le I + II + III + IV,$$

where $I = \sup_{t\in I_{\eta_k}} |\sum_{i=1}^{n} G_{k_i}^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k)(\tilde{z}_{k_i} - z_{k_i}^{(1)})|$, $II = \sup_{t\in I_{\eta_k}} |\sum_{i=1}^{n} G_{k_i}^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k)(z_{k_i}^{(1)} - z_{k_i}^{(2)})|$, $III = \sup_{t\in I_{\eta_k}} |\sum_{i=1}^{n} G_{k_i}^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k)(z_{k_i}^{(2)} - z_{k_i})|$, and $IV = \sup_{t\in I_{\eta_k}} |\hat{g}_k^{(\nu)}(t; \bar{\boldsymbol{\beta}}_k, (z_{k_i})_{i=1,\cdots,n}) - g_k^{(\nu)}(t)|$. By the Cauchy-Schwarz inequality and lemma 1 and lemma 2, under (M4) and the regularity conditions on smoothing, we have $I = O_p(\frac{1}{\sqrt{n} b_\nu^\nu b_0})$, $II = O_p(\delta_n b_\nu^\nu)$, $III = O_p(\gamma_{n\nu})$. Combining

$IV = O_p(\gamma_{n\nu})$ by theorem 1 of Chiou & Müller (1998) with the results for I, II and III, the proof is complete.

For the consistency result for the variance function, write

$$\hat{\sigma}^2(u) = \sum_{i=1}^{n} W_i(u; (\hat{\mu}_s)_{s=1,\ldots,n})(y_i - \hat{\mu}_i)^2 \tag{37}$$

where $W_i(u; (\hat{\mu}_s)_{s=1,\ldots,n})$ are the weight functions corresponding to the locally weighted least squares smoother with "design" points $\hat{\mu}_s = \sum_{k=1}^{M} \hat{g}_k(\boldsymbol{x}_s^T \hat{\boldsymbol{\beta}}_k)$.

Lemma 3

Given the weight function $W_i(u; (\hat{\mu}_s)_{s=1,\ldots,n})$ as in (37), if $\max_{1\le k\le M} \|\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = O_p(\frac{1}{\sqrt{n}})$, then under (M4) and the regularity conditions on smoothing,

$$\sup_{t\in I_\mu} \max_{1\le i\le n} |W_i(u; (\hat{\mu}_s)_{s=1,\ldots,n}) - W_i(u; (\mu_s)_{s=1,\ldots,n})| = O_p\left(\frac{1}{nb^2}(\gamma_{n0} + \delta_n)\right). \tag{38}$$

Proof of theorem 2. The result follows along the same lines as lemma 3 and theorem 2 in Chiou & Müller (1998).

By theorems 1 and 2 as well as the asymptotic distribution result in theorem 3 of Chiou & Müller (1998), we have for $\tilde{\boldsymbol{\beta}}_k \in \mathcal{B}_n^k$, $\sqrt{n}(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) \xrightarrow{\mathcal{D}} N(0, \Sigma_k^{-1})$. By Taylor expansion, $f(\tilde{\boldsymbol{\beta}}_k) = f(\boldsymbol{\beta}_k) + [(Df)(\boldsymbol{\beta}_k)](\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) + O(\|\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|^2)$. We note that $\sqrt{n}\|\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|^2 = o_p(1)$ and thus, by Slutsky's theorem, the remainder term can be ignored. We find that $\lim_{n\to\infty} \sqrt{n}(f(\tilde{\boldsymbol{\beta}}_k) - f(\boldsymbol{\beta}_k)) \overset{\mathcal{L}}{=} N(0, [(Df)(\boldsymbol{\beta}_k)]\Sigma_k^{-1}[(Df)(\boldsymbol{\beta}_k)]^T)$, where $\overset{\mathcal{L}}{=}$ denotes that the limiting distributions are the same.

We note that $\sup_{\tilde{}_k \in \mathcal{B}_n^k} |\hat{\boldsymbol{i}}_{\beta_k} - \boldsymbol{i}_{\beta_k}| = O_p(n(\alpha_n + (\gamma_{n0} + \delta_n)b^{-1})) = o_p(n)$ by theorems 1 and 2, and theorem A.1 of Chiou & Müller (1998) . It follows that under (M7) $(P_k - P_k^*)^T \tilde{\boldsymbol{\beta}}_k = o_p(\frac{1}{n})$, and accordingly, $\sqrt{n}(\tilde{\boldsymbol{\beta}}_k^* - \boldsymbol{\beta}_k^*) = (I_p - P_k)^T(\tilde{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) + o_p(\frac{1}{\sqrt{n}})$, along with (31), which implies the asymptotic result (32).

Table 1: *SSE, MSE, PSE (19) and PSE* (20) as a function of the number of indices m for orthogonality constrained QLUE implemented via projection with backfitting, for folate data*

m	$SSE(m)$	$MSE(m)$	$PSE(m)$	$PSE^*(m)$
1	143.60	1.1581	147.08	150.55
2	138.62	1.1179	145.32	152.03
3	137.93	1.1123	147.94	157.95

Table 2: *Orthogonality constrained QLUE implemented via projection with backfitting, for folate data ($M = 2$). For further details see text*

	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_2$
Predictor	Est.Coef. (s.e.)	Est.Coef. (s.e.)
FA	0.4856 (0.0095)	-0.1513 (0.0619)
BF	0.6181 (0.0052)	0.7560 (0.0188)
LF	0.6182 (0.0052)	-0.6369 (0.0324)
G.O.F.	$D = 131.85$	$D = 126.37$
	$P = 127.34$	$P = 124.11$
Bandwidth	62.16 (link)	189.71 (link)
	2.57 (var)	0.61 (var)

Table 3: *Seven treatments for the medfly reproduction experiment*

Treatment	Description of feeding schemes
1	Sugar only
2	Full Diet(Protein)
3	1:1, One day protein, one day sugar
4	1:3, One day protein, 3 days sugar
5	1:5, One day protein, 5 days sugar
6	1:10, One day protein, 10 days sugar
7	1:20, One day protein, 20 days sugar

Table 4: *Definition of variables for the medfly reproduction experiment*

Variables	Description of variables
y	Total number of eggs produced in lifetime as the dependent variable.
X_1	Lifetime of a female medfly (days).
$X_2, \ldots, X_7$	Indicator variables for the 7 treatments in Table 3.
	Treatment 1 is defined as the baseline by setting $X_i = 0$ for $i = 2, \ldots, 7$.
	For Treatment k, $2 \leq k \leq 7$, we set $X_k = 1$ and $X_i = 0$ for $i \neq k$ and $2 \leq i \leq 7$.

Table 5: *SSE, MSE and PSE as a function of m for constrained QLUE implemented via projection, with backfitting for the medfly reproduction data*

m	$SSE(m)$	$MSE(m)$	$PSE(m)$	$PSE^*(m)$
1	51,773,310	80,643	52,337,816	52,902,312
2	48,673,328	75,815	49,734,740	50,796,148
3	50,240,633	78,256	51,884,018	53,527,385
4	45,115,547	70,273	47,083,204	49,050,385
5	45,707,587	71,195	48,199,434	50,691,237
6	44,781,166	69,752	47,710,775	50,640,334

Table 6: *Orthogonality constrained QLUE implemented via projection with backfitting, for the medfly reproduction data ($M = 4$), including goodness-of-fit statistics and bandwidth choices*

	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_2$	$\hat{\boldsymbol{\beta}}_3$	$\hat{\boldsymbol{\beta}}_4$
Predictor	Est.Coef.(s.e.)	Est.Coef.(s.e.)	Est.Coef.(s.e.)	Est.Coef.(s.e.)
X_1	0.0058 (0.0003)	0.0186 (0.0012)	0.0036 (0.0002)	0.1045 (0.0283)
X_2	0.4289 (0.0128)	0.7313 (0.0223)	0.3584 (0.0114)	-.0739 (0.0948)
X_3	0.4463 (0.0136)	-.1790 (0.0572)	0.3492 (0.0162)	0.0668 (0.1954)
X_4	0.4722 (0.0136)	-.0623 (0.0657)	0.0676 (0.0311)	0.2953 (0.2220)
X_5	0.5084 (0.0136)	-.5710 (0.0325)	-.0555 (0.0160)	-.3972 (0.1218)
X_6	0.3388 (0.0153)	0.1290 (0.0176)	-.7339 (0.0070)	0.4868 (0.0829)
X_7	0.1434 (0.0147)	0.2938 (0.0178)	-.4510 (0.0071)	-.7052 (0.1108)
G.O.F.	$D = 698.0$	$D = 693.6$	$D = 616.4$	$D = 641.7$
	$P = 643.8$	$P = 623.9$	$P = 641.7$	$P = 643.0$
Bandwidth	0.2356 (link)	0.5962 (link)	0.2631 (link)	2.5060 (link)
	136.11 (var)	75.56 (var)	313.44 (var)	161.45 (var)

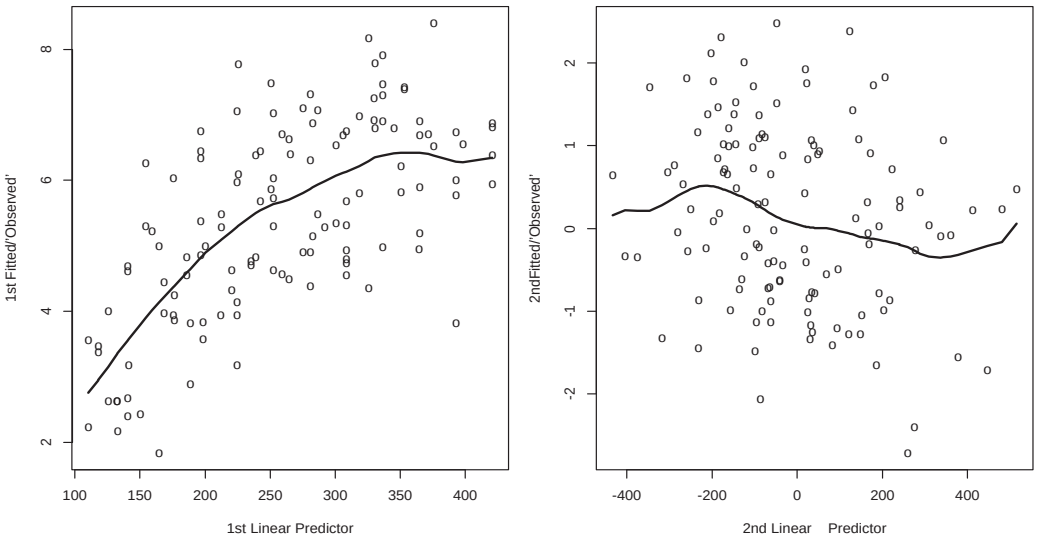

Figure 1: Link function estimates for first (left) and second (right) index for the folate bioassay data. Orthogonality constrained QLUE implemented via projection with backfitting, for the folate data (M=2).

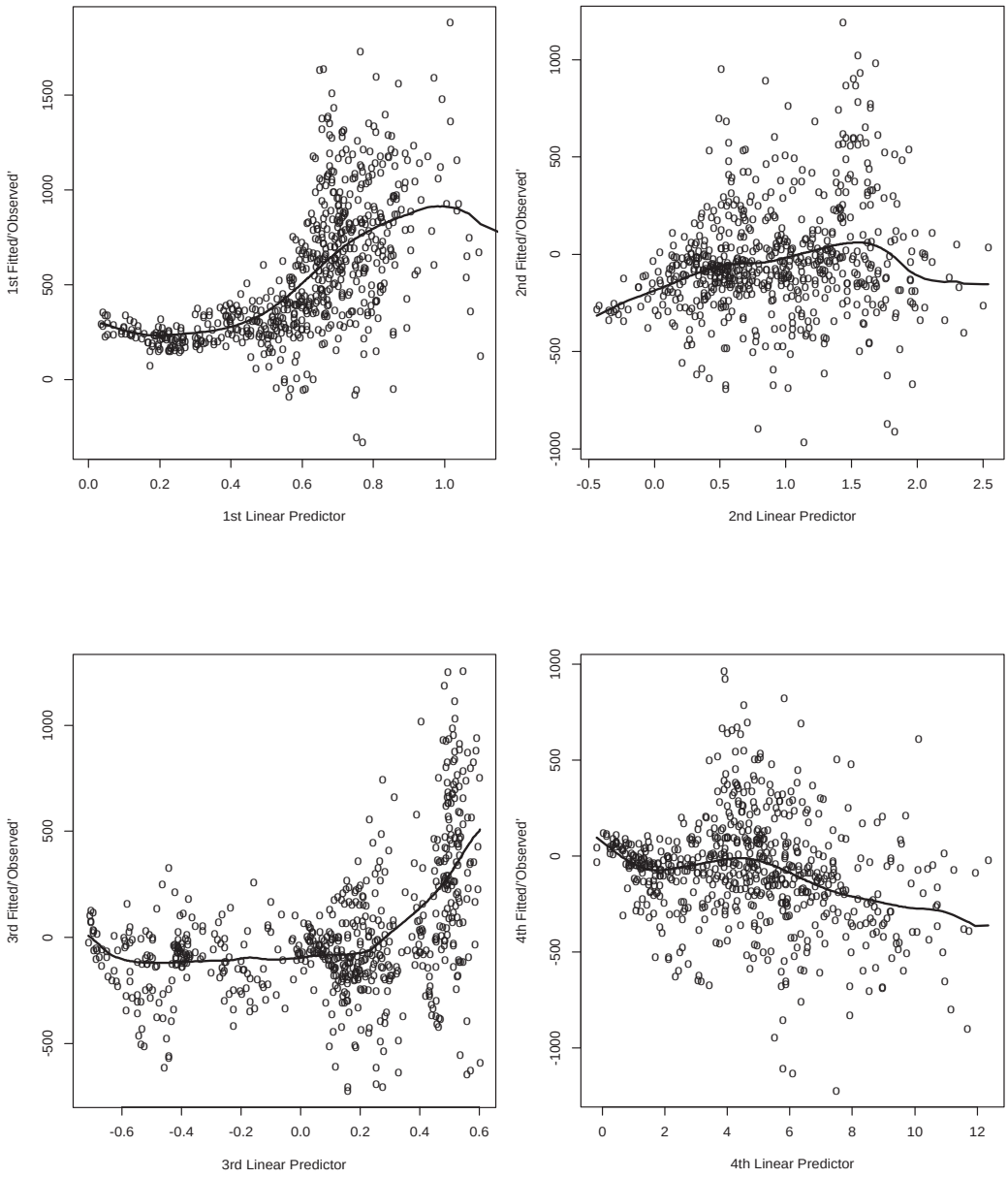

Figure 2: Link function estimates and associated scatterplots for the four indices used for fitting the medfly reproduction data. The fits are based on orthogonality constrained QLUE, implemented via projection with backfitting.

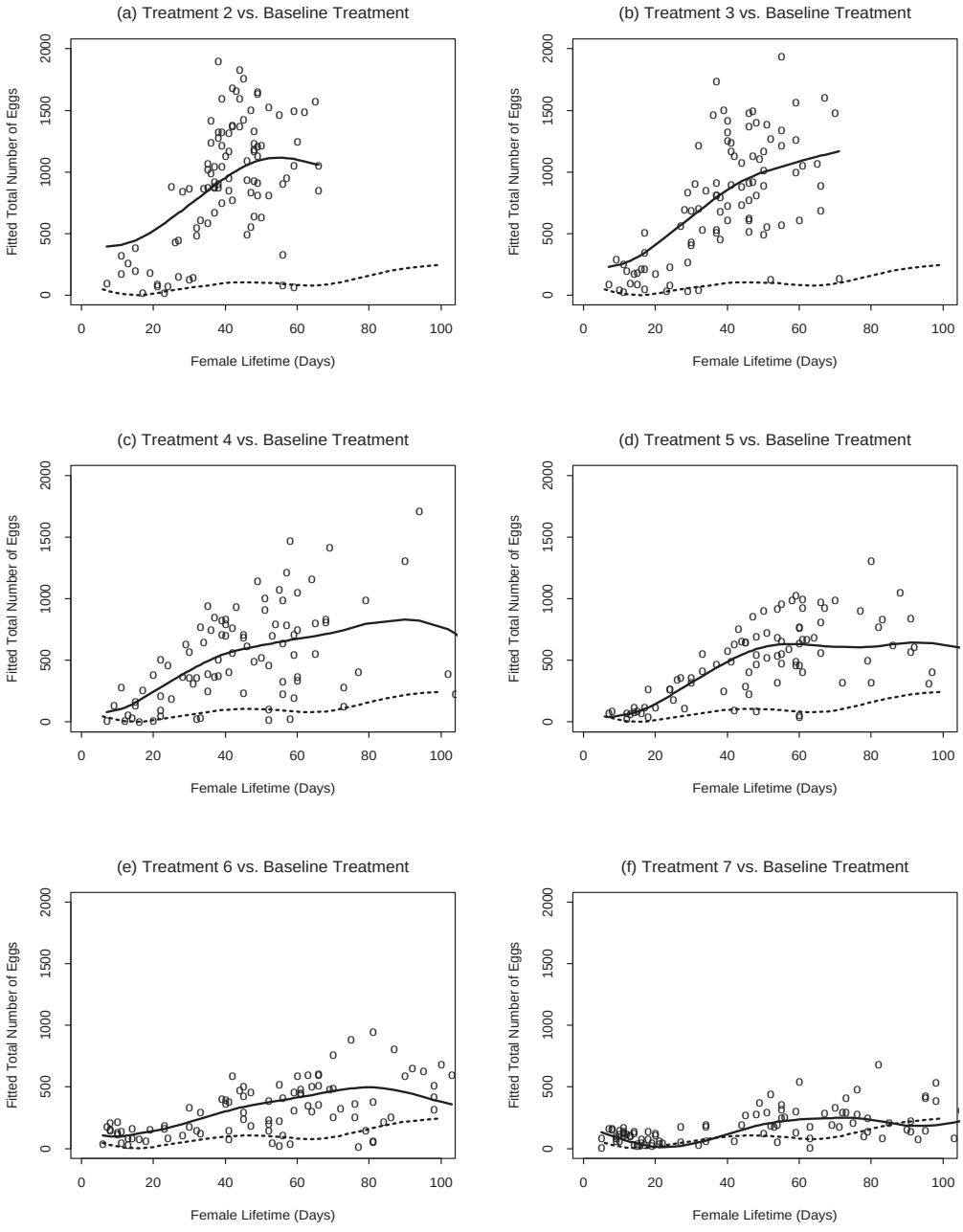

Figure 3: The relationship (solid curve) between total number of eggs and lifetime of the fly as implied by the fitted MUSE model, for each of the six protein treatments. Overlaid is the scatterplot of the raw data for each treatment group and the relation for the sugar only baseline treatment (dashed curve).