Generalized Additive Models with Flexible Response Functions
Supplementary Material

Elmar Spiegel[1], Thomas Kneib[1], Fabian Otto-Sobotka[2]
[1]University of Goettingen
[2]Carl von Ossietzky University Oldenburg

Contents

A Likelihood and Score Functions

In this section we list all used log-likelihood and score functions. Therefore we use the same notation as in the main part, but we always define the predictor to be semiparametric.

$$\eta_i := \boldsymbol{z}_i^\top \boldsymbol{\gamma}$$

Some vectors to shorten the notation:

$$\begin{aligned}
\boldsymbol{B}_\gamma &= (B_1(\boldsymbol{z}_i^\top \boldsymbol{\gamma}), \ldots, B_L(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \\
\boldsymbol{B}'_\gamma &= (B'_1(\boldsymbol{z}_i^\top \boldsymbol{\gamma}), \ldots, B'_L(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \\
\boldsymbol{B}''_\gamma &= (B''_1(\boldsymbol{z}_i^\top \boldsymbol{\gamma}), \ldots, B''_L(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \\
\boldsymbol{\nu} &= \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_L \end{pmatrix}
\end{aligned}$$

Furthermore, as in the paper we include the smoothing parameters in to their corresponding penalties: $\lambda \boldsymbol{K} = \boldsymbol{K}$

A.1 GAM

A.1.1 General GAM

Log-likelihood of the general GAM:

$$l(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \log(f(y_i|\boldsymbol{\gamma})) = \frac{y_i\theta_i - b(\theta_i)}{\phi}\omega_i - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma}$$

Score function of the general GAM:

$$\boldsymbol{s}(\boldsymbol{\gamma}) = \frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \frac{h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{\sigma_i^2}(y_i - h(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma}$$

Fisher Information of the general GAM:

$$\boldsymbol{H}(\boldsymbol{\gamma}) = -\frac{\partial \boldsymbol{s}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = -\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i \frac{1}{\sigma_i^2}\left(h''(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(y_i - h(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - (h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2\right) + \boldsymbol{K}_{\boldsymbol{\gamma}}$$

Expected Fisher Information of the general GAM:

$$\begin{aligned}
\boldsymbol{F}(\boldsymbol{\gamma}) = \mathbb{E}[\boldsymbol{H}(\boldsymbol{\gamma})] &= -\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i \frac{1}{\sigma_i^2}\left(h''(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\underbrace{\mathbb{E}\left[y_i - h(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\right]}_{=0} - (h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2 \right) + \boldsymbol{K}_{\boldsymbol{\gamma}} \\
&= \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^\top w_i + \boldsymbol{K}_{\boldsymbol{\gamma}}, \text{ with } w_i = \frac{(h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{\sigma_i^2}
\end{aligned}$$

Algorithm 1 (General IWLS)

$$
\begin{aligned}
\text{Start :}& \\
\boldsymbol{\gamma}^{(0)} &= (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{y} \\
\boldsymbol{\eta}^{(0)} &= \boldsymbol{Z} \boldsymbol{\gamma}^{(0)} \\
\text{Iterate } (k) \text{ :}& \\
\boldsymbol{\mu}^{(k)} &= h(\boldsymbol{\eta}^{(k-1)}) \\
\boldsymbol{y}^{(k)} &= \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{h'(\boldsymbol{\eta}^{(k-1)})} \\
\boldsymbol{w}^{(k)} &= \frac{\left(h'(\boldsymbol{\eta}^{(k-1)})\right)^2}{\sigma_i^2} \\
\boldsymbol{\gamma}^{(k)} &= (\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)} \\
\boldsymbol{\eta}^{(k)} &= \boldsymbol{Z} \boldsymbol{\gamma}^{(k)}
\end{aligned}
$$

A.2 Binomial Data

A.2.1 Logit Regression

Binomial log-likelihood:

$$l(\boldsymbol{\gamma}) = \sum_{i=1}^{n} y_i \log(\mu(\boldsymbol{z}_i)) + (1-y_i)\log(1-\mu(\boldsymbol{z}_i)) =$$
$$= \sum_{i=1}^{n} y_i \log\left(\frac{\mu(\boldsymbol{z}_i)}{1-\mu(\boldsymbol{z}_i)}\right) + \log(1-\mu(\boldsymbol{z}_i))$$

Canonical Logit-link:

$$g(\mu) = h^{-1}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) \Leftrightarrow \mu(\eta) = h(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}$$

Log-likelihood of the logit model:

$$l(\boldsymbol{\gamma}) = \sum_{i=1}^{n} y_i \boldsymbol{z}_i^\top \boldsymbol{\gamma} - \log(1+\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma}$$

Score function of the logit model:

$$\boldsymbol{s}(\boldsymbol{\gamma}) = \frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \left(y_i - \frac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{1+\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}\right) - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma}$$

Fisher Information of the logit model:

$$\boldsymbol{H}(\boldsymbol{\gamma}) = -\frac{\partial \boldsymbol{s}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i \frac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{(1+\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} + \boldsymbol{K}_{\boldsymbol{\gamma}}$$

Expected Fisher Information of the logit model:

$$\boldsymbol{F}(\boldsymbol{\gamma}) = \mathbb{E}[\boldsymbol{H}(\boldsymbol{\gamma})] = \boldsymbol{z}_i \boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}}, \text{ with } w_i = \frac{h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})^2}{h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} = h'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) = \mu(\boldsymbol{z}_i)(1-\mu(\boldsymbol{z}_i))$$

Algorithm 2 (Classical Logit Model)

Start :

$$\boldsymbol{\gamma}^{(0)} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{y}$$
$$\boldsymbol{\eta}^{(0)} = \boldsymbol{Z}\boldsymbol{\gamma}^{(0)}$$

Iterate (k) :

$$\boldsymbol{\mu}^{(k)} = \frac{\exp(\boldsymbol{\eta}^{(k-1)})}{1+\exp(\boldsymbol{\eta}^{(k-1)})}$$
$$\boldsymbol{y}^{(k)} = \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\boldsymbol{\mu}^{(k)}(1-\boldsymbol{\mu}^{(k)})}$$
$$\boldsymbol{w}^{(k)} = \boldsymbol{\mu}^{(k)}(1-\boldsymbol{\mu}^{(k)})$$
$$\boldsymbol{\gamma}^{(k)} = (\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)}$$
$$\boldsymbol{\eta}^{(k)} = \boldsymbol{Z}\boldsymbol{\gamma}^{(k)}$$

A.2.2 Indirect Estimation of the Response Function FlexGAM1

$$\mu_i = h(\Psi(\eta_i)) = \mu(\boldsymbol{z}_i)$$

where $h(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$ as before and Ψ is estimated as a monotonic P-spline with

$$\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\nu_l = \boldsymbol{B}_\gamma \boldsymbol{\nu}$$

Log-likelihood of FlexGAM1 for binomial data:

$$\begin{aligned} l(\boldsymbol{\gamma}, \Psi) &= \sum_{i=1}^{n} y_i \Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) - \log(1 + \exp(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_\gamma \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} y_i \boldsymbol{B}_\gamma \boldsymbol{\nu} - \log(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_\gamma \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Score functions of FlexGAM1 for binomial data:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}} &= \sum_{i=1}^{n} \left(y_i - \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})} \right) \boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i - \boldsymbol{K}_\gamma \boldsymbol{\gamma} \\ &= \sum_{i=1}^{n} (y_i - h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma})))\Psi'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\boldsymbol{z}_i - \boldsymbol{K}_\gamma \boldsymbol{\gamma} \\ \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} &= \sum_{i=1}^{n} \left(y_i - \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})} \right) \boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} \left(y_i - h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \right) \boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Fisher Information of FlexGAM1 for binomial data:

$$\begin{aligned} -\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}\boldsymbol{\gamma}} &= -\sum_{i=1}^{n} -\frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})(\boldsymbol{B}'_\gamma \boldsymbol{\nu})^2 \boldsymbol{z}_i \boldsymbol{z}_i}{(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}))^2} + \left(y_i - \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})} \right) \boldsymbol{B}''_\gamma \boldsymbol{\nu} \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_\gamma \\ -\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}\boldsymbol{\nu}} &= \sum_{i=1}^{n} \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}) \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top}{(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}))^2} + \boldsymbol{K}_\nu \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} &= \sum_{i=1}^{n} \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}))^2} \boldsymbol{B}_\gamma^\top \boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i + \left(y_i - \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})} \right) \boldsymbol{B}_\gamma^{\prime\top} \boldsymbol{z}_i \end{aligned}$$

Expected Fisher Information of FlexGAM1 for binomial data:

$$\begin{aligned} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}\boldsymbol{\gamma}} \right] &= \sum_{i=1}^{n} \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}))^2} (\boldsymbol{B}'_\gamma \boldsymbol{\nu})^2 \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_\gamma \\ &= \sum_{i=1}^{n} h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))(1 - h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma})))(\Psi'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2 \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_\gamma \\ F_{\boldsymbol{\nu},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}\boldsymbol{\nu}} \right] &= \sum_{i=1}^{n} \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}))^2} \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ &= \sum_{i=1}^{n} h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))(1 - h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))) \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} \right] &= \sum_{i=1}^{n} \frac{\exp(\boldsymbol{B}_\gamma \boldsymbol{\nu})}{(1 + \exp(\boldsymbol{B}_\gamma \boldsymbol{\nu}))^2} \boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \\ &= \sum_{i=1}^{n} h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))(1 - h(\Psi(\boldsymbol{z}_i^\top \boldsymbol{\gamma})))\Psi'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \\ F(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \begin{pmatrix} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} & F_{\boldsymbol{\gamma},\boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu},\boldsymbol{\nu}} \end{pmatrix} \end{aligned}$$

In the estimation procedure we fix Ψ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = \mu(\boldsymbol{z}_i)(1-\mu(\boldsymbol{z}_i))(\Psi'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2$$

Working responses:

$$\begin{aligned}
y_i^{(k)} &= \eta_i^{(k-1)} + \frac{y_i - \mu_i^{(k)}}{\frac{\partial h(\eta_i^{(k-1)})}{\partial \eta}} \\
&= \eta_i^{(k-1)} + \frac{y_i - h(\Psi(\eta_i^{(k-1)}))}{h'(\Psi(\eta_i^{(k-1)}))\Psi'(\eta_i^{(k-1)})} \\
&= \eta_i^{(k-1)} + \frac{y_i - h(\Psi(\eta_i^{(k-1)}))}{h(\Psi(\eta_i^{(k-1)}))(1-h(\Psi(\eta_i^{(k-1)})))\Psi'(\eta_i^{(k-1)})} \\
&= \eta_i^{(k-1)} + \frac{y_i - \boldsymbol{\mu}^{(k)}}{\boldsymbol{\mu}^{(k)}(1-\boldsymbol{\mu}^{(k)})\Psi'(\eta_i^{(k-1)})}
\end{aligned}$$

Algorithm 3 (FlexGAM1 Binomial)

Start:

$$\hat{\boldsymbol{\gamma}}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots\texttt{, family=binomial(link=logit))}$$

$$\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}$$

Outer (m) :

$$\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}) = \texttt{scam(}\boldsymbol{y} \sim s(\boldsymbol{\eta}^{(k-1)}\texttt{, bs="mpi"), family=binomial(link=logit))}$$

Inner (k) :

$$\begin{aligned}
\boldsymbol{\mu}^{(k)} &= h(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)})) = \frac{\exp(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}))}{1+\exp(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}))} \\
\boldsymbol{y}^{(k)} &= \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\boldsymbol{\mu}^{(k)}(1-\boldsymbol{\mu}^{(k)})\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})} \\
\boldsymbol{w}^{(k)} &= \boldsymbol{\mu}^{(k)}(1-\boldsymbol{\mu}^{(k)})\left(\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2 \\
\hat{\boldsymbol{\gamma}}^{(k)} &= \left(\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_\gamma\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)} \\
&= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)} \\
\boldsymbol{\eta}_1^{(k)} &= \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}
\end{aligned}$$

A.2.3 Direct Estimation of the Response Function FlexGAM2

$$\mu_i = \mathbb{E}[y_i|\boldsymbol{z}_i] = \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}),$$

with

$$\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\nu_l = \boldsymbol{B}_\gamma \boldsymbol{\nu}$$

Log-likelihood of FlexGAM2 for binomial data:

$$\begin{aligned} l(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \sum_{i=1}^{n} y_i \log(\mu(\boldsymbol{z}_i)) + (1 - y_i)\log(1 - \mu(\boldsymbol{z}_i)) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} y_i \log\left(\frac{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}\right) + \log(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} y_i \log\left(\frac{\boldsymbol{B}_\gamma \boldsymbol{\nu}}{1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}}\right) + \log\left(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}\right) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Score function of FlexGAM2 for binomial data:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}} &= \sum_{i=1}^{n} \frac{\boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})}(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} \\ &= \sum_{i=1}^{n} \frac{\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) \boldsymbol{z}_i}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))}(y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} \\ \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} &= \sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma^\top}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})}(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) - \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma^\top}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))}(y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Fisher Information of FlexGAM2 for binomial data:

$$\begin{aligned} -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \boldsymbol{B}''_\gamma \boldsymbol{\nu} \boldsymbol{z}_i \boldsymbol{z}_i - (\boldsymbol{B}'_\gamma \boldsymbol{\nu})^2 (1 - 2\boldsymbol{B}_\gamma \boldsymbol{\nu}) \boldsymbol{z}_i \boldsymbol{z}_i}{(\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}))^2}(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) - \frac{(\boldsymbol{B}'_\gamma \boldsymbol{\nu})^2 \boldsymbol{z}_i \boldsymbol{z}_i}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})} + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} -\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top \frac{(1 - 2\boldsymbol{B}_\gamma \boldsymbol{\nu})}{(\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}))^2}(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) - \frac{1}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})} \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \boldsymbol{B}_\gamma^{'\top} \boldsymbol{z}_i - \boldsymbol{B}_\gamma^\top \boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i (1 - 2\boldsymbol{B}_\gamma \boldsymbol{\nu})}{(\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu}))^2}(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) - \frac{\boldsymbol{B}_\gamma^\top \boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})} \end{aligned}$$

Expected Fisher Information of FlexGAM2 for binomial data:

$$\begin{aligned} F_{\boldsymbol{\gamma}, \gamma} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \boldsymbol{\gamma}}\right] &= \sum_{i=1}^{n} \frac{1}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})} (\boldsymbol{B}'_\gamma \boldsymbol{\nu})^2 \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ &= \sum_{i=1}^{n} \frac{1}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))} \left(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\right)^2 \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ F_{\boldsymbol{\nu}, \boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \boldsymbol{\nu}}\right] &= \sum_{i=1}^{n} \frac{1}{\boldsymbol{B}_\gamma \boldsymbol{\nu}(1 - \boldsymbol{B}_\gamma \boldsymbol{\nu})} \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ &= \sum_{i=1}^{n} \frac{1}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))} \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ F_{\boldsymbol{\gamma}, \boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}}\right] &= \sum_{i=1}^{n} \frac{1}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))} \hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \\ F(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \begin{pmatrix} F_{\boldsymbol{\gamma}, \gamma} & F_{\boldsymbol{\gamma}, \boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma}, \boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu}, \boldsymbol{\nu}} \end{pmatrix} \end{aligned}$$

In the estimation procedure we fix $\hat{h}$ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^\top w_i + \boldsymbol{K}_\gamma, \qquad \text{with } w_i = \frac{(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})(1 - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))}$$

Algorithm 4 (FlexGAM2 Binomial)

Start:

$$\hat{\boldsymbol{\gamma}}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, family=binomial(link=logit))}$$

$$\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}$$

Outer (m) :

$$\hat{\boldsymbol{\nu}}^{(m)} = \text{argmin}\left(\sum_{i=1}^{n}\left(y_i - \sum_{l=1}^{L} B_l(\eta_i^{(k-1)})\nu_l\right)^2 + \boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu}\right) \Rightarrow \hat{h}^{(m)}$$

Inner (k) :

$$\boldsymbol{\mu}^{(k)} = \hat{h}^{(m)}(\boldsymbol{\eta}^{(k-1)})$$

$$\boldsymbol{y}^{(k)} = \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})}$$

$$\boldsymbol{w}^{(k)} = \frac{\left(\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2}{\boldsymbol{\mu}^{(k)}(1 - \boldsymbol{\mu}^{(k)})}$$

$$\hat{\boldsymbol{\gamma}}^{(k)} = \left(\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_\gamma\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)}$$

$$= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)}$$

$$\boldsymbol{\eta}_1^{(k)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}$$

In the binomial case the constraints for the estimation of $\hat{h}$ are that the spline has to be strictly monotonic increasing and the fitted values have to be positive and smaller than 1. So the coefficients $\boldsymbol{\nu}$ have to fulfill:

$$0 < \nu_l < \nu_{l+1} < 1$$

A.3 Poisson Data

$$f(y_i|\boldsymbol{\gamma}) = \frac{\lambda^{y_i}\exp(-\lambda)}{y_i!}$$

Likelihood of Poisson data:

$$\begin{aligned}
L(\boldsymbol{\gamma}|\boldsymbol{y}) &= \prod_{i=1}^{n} \frac{h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})^{y_i}\exp(-h(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))}{y_i!} \\
l(\boldsymbol{\gamma}|\boldsymbol{y}) &= \sum_{i=1}^{n} y_i \log(h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})) - h(\boldsymbol{z}_i^\top\boldsymbol{\gamma}) - \log(y_i!)
\end{aligned}$$

A.3.1 Indirect estimation of the Response Function FlexGAM1

$$\mu_i = \mathbb{E}[y_i] = \lambda(\boldsymbol{z}_i) = h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})) = \exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))$$

with

$$\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\nu_l = \boldsymbol{B}_\gamma\boldsymbol{\nu}$$

Log-likelihood of FlexGAM1 for Poisson data:

$$\begin{aligned}
l(\boldsymbol{\gamma},\boldsymbol{\nu}) &= \sum_{i=1}^{n} y_i\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}) - \exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})) - \log(y_i!) - \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{K}_\nu\boldsymbol{\nu} \\
&= \sum_{i=1}^{n} y_i\boldsymbol{B}_\gamma\boldsymbol{\nu} - \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}) - \log(y_i!) - \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{K}_\nu\boldsymbol{\nu}
\end{aligned}$$

Score function of FlexGAM1 for Poisson data

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}} &= \sum_{i=1}^{n} (y_i - \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}))\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{z}_i - \boldsymbol{K}_\gamma\gamma \\
&= \sum_{i=1}^{n} (y_i - h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})))\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma} \\
\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}} &= \sum_{i=1}^{n} (y_i - \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}))\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu\boldsymbol{\nu} \\
&= \sum_{i=1}^{n} (y_i - h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})))\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu\boldsymbol{\nu}
\end{aligned}$$

Fisher Information of FlexGAM1 for Poisson data:

$$\begin{aligned}
-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \left((y_i - \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}))\boldsymbol{B}''_\gamma\boldsymbol{\nu} - \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})(\boldsymbol{B}'_\gamma\boldsymbol{\nu})^2\right)\boldsymbol{z}_i\boldsymbol{z}_i + \boldsymbol{K}_\gamma \\
-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}} &= -\sum_{i=1}^{n} -\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma^\top\boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\
-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\nu}} &= -\sum_{i=1}^{n} (-\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma^\top)\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{z}_i + (y_i - \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}))\boldsymbol{B}_\gamma^{\prime\top}\boldsymbol{z}_i
\end{aligned}$$

Expected Fisher Information of FlexGAM1 for Poisson data:

$$
\begin{aligned}
F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\boldsymbol{\gamma}}\right] &= \sum_{i=1}^{n} \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})(\boldsymbol{B}_\gamma'\boldsymbol{\nu})^2 \boldsymbol{z}_i\boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\
&= \sum_{i=1}^{n} h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))(\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2 \boldsymbol{z}_i\boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\
F_{\boldsymbol{\nu},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}\boldsymbol{\nu}}\right] &= \sum_{i=1}^{n} \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma^\top\boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\
&= \sum_{i=1}^{n} h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\boldsymbol{B}_\gamma^\top\boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\
F_{\boldsymbol{\gamma},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\nu}}\right] &= \sum_{i=1}^{n} \exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma'\boldsymbol{\nu}\boldsymbol{B}_\gamma^\top\boldsymbol{z}_i \\
&= \sum_{i=1}^{n} h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\boldsymbol{B}_\gamma^\top\boldsymbol{z}_i \\
F(\boldsymbol{\gamma},\boldsymbol{\nu}) &= \begin{pmatrix} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} & F_{\boldsymbol{\gamma},\boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu},\boldsymbol{\nu}} \end{pmatrix}
\end{aligned}
$$

In the estimation procedure we fix Ψ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$
\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i\boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = \exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))(\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2
$$

Algorithm 5 (FlexGAM1 Poisson)

Start:

$$
\hat{\boldsymbol{\gamma}}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, family=poisson(link=log))}
$$

$$
\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}
$$

Outer (m) :

$$
\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}) = \texttt{scam(}\boldsymbol{y} \sim s(\boldsymbol{\eta}^{(k-1)}\texttt{, bs="mpi"), family=poisson(link=log))}
$$

Inner (k) :

$$
\begin{aligned}
\boldsymbol{\mu}^{(k)} &= h(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)})) = \exp(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)})) \\
\boldsymbol{y}^{(k)} &= \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\boldsymbol{\mu}^{(k)}\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})} \\
\boldsymbol{w}^{(k)} &= \boldsymbol{\mu}^{(k)}\left(\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2 \\
\hat{\boldsymbol{\gamma}}^{(k)} &= \left(\boldsymbol{Z}^\top\boldsymbol{W}^{(k)}\boldsymbol{Z} + \boldsymbol{K}_{\boldsymbol{\gamma}}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{W}^{(k)}\boldsymbol{y}^{(k)} \\
&= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)} \\
\boldsymbol{\eta}_1^{(k)} &= \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}
\end{aligned}
$$

A.3.2 Direct estimation of the Response Function FlexGAM2

$$\mu_i = \mathbb{E}[y_i] = \lambda(\boldsymbol{z}_i) = \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})$$

with

$$\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\nu_l = \boldsymbol{B}_\gamma \boldsymbol{\nu}$$

Log-likelihood of FlexGAM2 for Poisson data:

$$\begin{aligned} l(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \sum_{i=1}^{n} y_i \log(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) - \log(y_i!) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} y_i \log(\boldsymbol{B}_\gamma \boldsymbol{\nu}) - \boldsymbol{B}_\gamma \boldsymbol{\nu} - \log(y_i!) - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Score function of FlexGAM2 for Poisson data:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}} &= \sum_{i=1}^{n} (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \frac{1}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} \boldsymbol{B}_\gamma' \boldsymbol{\nu} \boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} \\ &= \sum_{i=1}^{n} (y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \frac{1}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} \hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) \boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} \\ \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} &= \sum_{i=1}^{n} (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \frac{1}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} \boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} (y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \frac{1}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} \boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Fisher Information of FlexGAM2 for Poisson data:

$$\begin{aligned} -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \left((y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \frac{\boldsymbol{B}_\gamma'' \boldsymbol{\nu}}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} - \frac{(\boldsymbol{B}_\gamma' \boldsymbol{\nu})^2}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} \right) \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} -\frac{\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} - (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \frac{\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} + \boldsymbol{K}_\nu \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) \frac{\boldsymbol{B}_\gamma'^\top \boldsymbol{z}_i}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} - y_i \frac{\boldsymbol{B}_\gamma' \boldsymbol{\nu} \boldsymbol{z}_i \boldsymbol{B}_\gamma^\top}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} \end{aligned}$$

Expected Fisher Information of FlexGAM2 for Poisson data:

$$\begin{aligned} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \boldsymbol{\gamma}} \right] &= \sum_{i=1}^{n} \frac{(\boldsymbol{B}_\gamma' \boldsymbol{\nu})^2}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ &= \sum_{i=1}^{n} \frac{(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ F_{\boldsymbol{\nu},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \boldsymbol{\nu}} \right] &= \sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} + \boldsymbol{K}_\nu \\ &= \sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} + \boldsymbol{K}_\nu \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} \right] &= \sum_{i=1}^{n} \frac{\boldsymbol{B}_\gamma' \boldsymbol{\nu} \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} \\ &= \sum_{i=1}^{n} \frac{\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \end{aligned}$$

$$F(\boldsymbol{\gamma}, \boldsymbol{\nu}) = \begin{pmatrix} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} & F_{\boldsymbol{\gamma},\boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu},\boldsymbol{\nu}} \end{pmatrix}$$

In the estimation procedure we fix $\hat{h}$ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$\boldsymbol{F}(\gamma) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = \frac{\left(\hat{h}'(\boldsymbol{z}_i^\top \gamma)\right)^2}{\hat{h}(\boldsymbol{z}_i^\top \gamma)}$$

Algorithm 6 (FlexGAM2 Poisson)

Start:

$$\hat{\gamma}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots \texttt{, family=poisson(link=log))}$$

$$\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\gamma}^{(0)} \quad \Rightarrow \quad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}$$

Outer (m) :

$$\hat{\boldsymbol{\nu}}^{(m)} = \text{argmin}\left(\sum_{i=1}^{n}\left(y_i - \sum_{l=1}^{L} B_l(\eta_i^{(k-1)})\nu_l\right)^2 + \boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu}\right) \Rightarrow \hat{h}^{(m)}$$

Inner (k) :

$$\boldsymbol{\mu}^{(k)} = \hat{h}^{(m)}(\boldsymbol{\eta}^{(k-1)})$$

$$\boldsymbol{y}^{(k)} = \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})}$$

$$\boldsymbol{w}^{(k)} = \frac{\left(\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2}{\boldsymbol{\mu}^{(k)}}$$

$$\hat{\gamma}^{(k)} = \left(\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_{\boldsymbol{\gamma}}\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)}$$

$$= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots \texttt{, weight = } \boldsymbol{w}^{(k)}\texttt{, family=gaussian)}$$

$$\boldsymbol{\eta}_1^{(k)} = \boldsymbol{Z}\hat{\gamma}^{(k)} \quad \Rightarrow \quad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}$$

However in the Poisson case the constraints for the estimation of $\hat{h}$ are that the spline has to be strictly monotone increasing and the fitted values have to be positive. There is no upper bound for the fitted values and the spline. So the coefficients $\boldsymbol{\nu}$ have to fulfill:

$$0 < \nu_l < \nu_{l+1}$$

A.4 Gaussian Data

$$f(y_i|\boldsymbol{\gamma},\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - h(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2}{2\sigma^2}\right)$$

Likelihood of Gaussian data:

$$L(\boldsymbol{\gamma}|\boldsymbol{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - h(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2}{2\sigma^2}\right)$$
$$l(\boldsymbol{\gamma}|\boldsymbol{y}) = \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_i - h(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2}{2\sigma^2}$$

A.4.1 Indirect Estimation of the Response Function FlexGAM1

$$\mu_i = \mathbb{E}[y_i] = h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})) = \Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})$$

with

$$\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\nu_l = \boldsymbol{B}_\gamma\boldsymbol{\nu}$$

Log-likelihood of FlexGAM1 for Gaussian data:

$$\begin{aligned}
l(\boldsymbol{\gamma},\boldsymbol{\nu}) &= \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_i - \Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2}{2\sigma^2} - \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{K}_\gamma\boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{K}_\nu\boldsymbol{\nu} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_i - \boldsymbol{B}_\gamma\boldsymbol{\nu})^2}{2\sigma^2} - \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{K}_\gamma\boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{K}_\nu\boldsymbol{\nu}
\end{aligned}$$

Score function of FlexGAM1 for Gaussian data

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}} &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y_i - \Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\boldsymbol{z}_i - \boldsymbol{K}_\gamma\boldsymbol{\gamma} \\
&= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y_i - \boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{z}_i - \boldsymbol{K}_\gamma\boldsymbol{\gamma} \\
\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}} &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y - \Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu\boldsymbol{\nu} \\
&= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y - \boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu\boldsymbol{\nu}
\end{aligned}$$

Fisher Information of FlexGAM1 for Gaussian data:

$$\begin{aligned}
-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \frac{1}{\sigma^2}\left((-1)\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{B}'_\gamma\boldsymbol{\nu} + (y - \boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}''_\gamma\boldsymbol{\nu}\right)\boldsymbol{z}_i\boldsymbol{z}_i + \boldsymbol{K}_\gamma \\
-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}} &= -\sum_{i=1}^{n} -\frac{1}{\sigma^2}\boldsymbol{B}_\gamma^\top\boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\
-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\nu}} &= -\sum_{i=1}^{n} \frac{1}{\sigma^2}((-\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{z}_i)\boldsymbol{B}_\gamma^\top + (y_i - \boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma^{\prime\top}\boldsymbol{z}_i)
\end{aligned}$$

Expected Fisher Information of FlexGAM1 for Gaussian data:

$$
\begin{aligned}
F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\boldsymbol{\gamma}}\right] &= \sum_{i=1}^{n}\frac{1}{\sigma^2}(\boldsymbol{B}'_{\gamma}\boldsymbol{\nu})^2\boldsymbol{z}_i\boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\
&= \sum_{i=1}^{n}\frac{1}{\sigma^2}(\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2\boldsymbol{z}_i\boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\
F_{\boldsymbol{\nu},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}\boldsymbol{\nu}}\right] &= \sum_{i=1}^{n}\frac{1}{\sigma^2}\boldsymbol{B}_{\gamma}^\top\boldsymbol{B}_{\gamma}^\top + \boldsymbol{K}_{\nu} \\
F_{\boldsymbol{\gamma},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\nu}}\right] &= \sum_{i=1}^{n}\frac{1}{\sigma^2}\boldsymbol{B}'_{\gamma}\boldsymbol{\nu}\boldsymbol{B}_{\gamma}^\top\boldsymbol{z}_i \\
&= \sum_{i=1}^{n}\frac{1}{\sigma^2}\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\boldsymbol{B}_{\gamma}^\top\boldsymbol{z}_i \\
F(\boldsymbol{\gamma},\boldsymbol{\nu}) &= \begin{pmatrix} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} & F_{\boldsymbol{\gamma},\boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu},\boldsymbol{\nu}} \end{pmatrix}
\end{aligned}
$$

In the estimation procedure we fix Ψ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$
\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = (\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2
$$

The variance σ^2 is not necessary for the estimation part. For calculating the standard errors we use

$$
\hat{\sigma}^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{n-p},
$$

where p are the estimated degrees of freedom of the covariate effects.

Algorithm 7 (FlexGAM1 Gaussian)

Start:

$$
\hat{\boldsymbol{\gamma}}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, family=gaussian())}
$$

$$
\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}
$$

Outer (m) :

$$
\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}) = \texttt{scam(}\boldsymbol{y} \sim s(\boldsymbol{\eta}^{(k-1)}\texttt{, bs="mpi"), family=gaussian())}
$$

Inner (k) :

$$
\begin{aligned}
\boldsymbol{\mu}^{(k)} &= h(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)})) = \Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}) \\
\boldsymbol{y}^{(k)} &= \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})} \\
\boldsymbol{w}^{(k)} &= \left(\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2 \\
\hat{\boldsymbol{\gamma}}^{(k)} &= \left(\boldsymbol{Z}^\top\boldsymbol{W}^{(k)}\boldsymbol{Z} + \boldsymbol{K}_{\boldsymbol{\gamma}}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{W}^{(k)}\boldsymbol{y}^{(k)} \\
&= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)} \\
\boldsymbol{\eta}_1^{(k)} &= \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}
\end{aligned}
$$

A.4.2 Direct Estimation of the Response Function FlexGAM2

$$\mu_i = \mathbb{E}[y_i] = \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})$$

with

$$\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top \gamma)\nu_l = \boldsymbol{B}_\gamma \boldsymbol{\nu}$$

Log-likelihood of FlexGAM2 for Gaussian data:

$$\begin{aligned} l(\gamma, \boldsymbol{\nu}) &= \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_i - \hat{h}(\boldsymbol{z}_i^\top \gamma))^2}{2\sigma^2} - \frac{1}{2}\gamma^\top \boldsymbol{K}_\gamma \gamma - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu})^2}{2\sigma^2} - \frac{1}{2}\gamma^\top \boldsymbol{K}_\gamma \gamma - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Score function of FlexGAM2 for Gaussian data

$$\begin{aligned} \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}} &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y_i - \hat{h}(\boldsymbol{z}_i^\top \gamma))\hat{h}'(\boldsymbol{z}_i^\top \gamma)\boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}}\gamma \\ &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu})\boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}}\gamma \\ \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y - \hat{h}(\boldsymbol{z}_i^\top \gamma))\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(y - \boldsymbol{B}_\gamma \boldsymbol{\nu})\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Fisher Information of FlexGAM2 for Gaussian data:

$$\begin{aligned} -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \frac{1}{\sigma^2}\left((-1)\boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{B}'_\gamma \boldsymbol{\nu} + (y - \boldsymbol{B}_\gamma \boldsymbol{\nu})\boldsymbol{B}''_\gamma \boldsymbol{\nu}\right)\boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} -\frac{1}{\sigma^2}\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} \frac{1}{\sigma^2}((-\boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{z}_i)\boldsymbol{B}_\gamma^\top + (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu})\boldsymbol{B}_\gamma'^\top \boldsymbol{z}_i) \end{aligned}$$

Expected Fisher Information of FlexGAM2 for Gaussian data:

$$\begin{aligned} F_{\boldsymbol{\gamma},\gamma} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}\boldsymbol{\gamma}}\right] &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(\boldsymbol{B}'_\gamma \boldsymbol{\nu})^2 \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ &= \sum_{i=1}^{n} \frac{1}{\sigma^2}(\hat{h}'(\boldsymbol{z}_i^\top \gamma))^2 \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ F_{\boldsymbol{\nu},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}\boldsymbol{\nu}}\right] &= \sum_{i=1}^{n} \frac{1}{\sigma^2}\boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}}\right] &= \sum_{i=1}^{n} \frac{1}{\sigma^2}\boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \\ &= \sum_{i=1}^{n} \frac{1}{\sigma^2}\hat{h}'(\boldsymbol{z}_i^\top \gamma)\boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \\ F(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \begin{pmatrix} F_{\boldsymbol{\gamma},\gamma} & F_{\boldsymbol{\gamma},\boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu},\boldsymbol{\nu}} \end{pmatrix} \end{aligned}$$

In the estimation procedure we fix $\hat{h}$ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = (\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2$$

The variance σ^2 is not necessary for the estimation part. For calculating the standard errors we use

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{n-p},$$

where p are the estimated degrees of freedom of the covariate effects.

Algorithm 8 (FlexGAM2 Gaussian)

Start:

$$\hat{\gamma}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots\texttt{, family=gaussian())}$$

$$\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\gamma}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}$$

Outer (m) :

$$\hat{\boldsymbol{\nu}}^{(m)} = \text{argmin} \left(\sum_{i=1}^{n} \left(y_i - \sum_{l=1}^{L} B_l(\eta_i^{(k-1)}) \nu_l \right)^2 + \boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \right) \Rightarrow \hat{h}^{(m)}$$

Inner (k) :

$$\boldsymbol{\mu}^{(k)} = \hat{h}^{(m)}(\boldsymbol{\eta}^{(k-1)})$$

$$\boldsymbol{y}^{(k)} = \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})}$$

$$\boldsymbol{w}^{(k)} = \left(\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2$$

$$\hat{\gamma}^{(k)} = \left(\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_\gamma\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)}$$

$$= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)}$$

$$\boldsymbol{\eta}_1^{(k)} = \boldsymbol{Z}\hat{\gamma}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}$$

However in the Gaussian case the constraints for the estimation of $\hat{h}$ are that the spline has to be strictly monotone increasing. There is no upper or lower bound for the fitted value and the spline. So the coefficients $\boldsymbol{\nu}$ have to fulfill:

$$\nu_l < \nu_{l+1}$$

A.5 Gamma Data

$$f(y_i|\gamma,\kappa)=\frac{1}{\Gamma(\kappa)}\left(\frac{\kappa}{h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})}\right)^\kappa y_i^{k-1}\exp\left(-\frac{\kappa y_i}{h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})}\right)$$

Likelihood of Gamma data:

$$L(\boldsymbol{\gamma}|\boldsymbol{y})=\prod_{i=1}^n\frac{1}{\Gamma(\kappa)}\left(\frac{\kappa}{h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})}\right)^\kappa y_i^{k-1}\exp\left(-\frac{\kappa y_i}{h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})}\right)$$

$$l(\boldsymbol{\gamma}|\boldsymbol{y})=\sum_{i=1}^n-\log(\Gamma(\kappa))+\kappa\log(\kappa)-\kappa\log(h(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))+(\kappa-1)\log(y_i)-\frac{\kappa y_i}{h(\boldsymbol{z}_i^\top\boldsymbol{\gamma})}$$

A.5.1 Indirect Estimation of the Response Function FlexGAM1

$$\mu_i=\mathbb{E}[y_i]=h(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))=\exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))$$

with

$$\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})=\sum_{l=1}^L B_l(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\nu_l=\boldsymbol{B}_\gamma\boldsymbol{\nu}$$

Log-likelihood of FlexGAM1 for Gamma data:

$$\begin{aligned}l(\boldsymbol{\gamma},\boldsymbol{\nu})&=\sum_{i=1}^n-\log(\Gamma(\kappa))+\kappa\log(\kappa)-\kappa\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})+(\kappa-1)\log(y_i)-\kappa y_i\exp(-\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))-\frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma}-\frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{K}_\nu\boldsymbol{\nu}\\&=\sum_{i=1}^n-\log(\Gamma(\kappa))+\kappa\log(\kappa)-\kappa\boldsymbol{B}_\gamma\boldsymbol{\nu}+(\kappa-1)\log(y_i)-\kappa y_i\exp(-\boldsymbol{B}_\gamma\boldsymbol{\nu})-\frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma}-\frac{1}{2}\boldsymbol{\nu}^\top\boldsymbol{K}_\nu\boldsymbol{\nu}\end{aligned}$$

Score function of FlexGAM1 for Gamma data

$$\begin{aligned}\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}}&=\sum_{i=1}^n\kappa\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma})\exp(-\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\left[y_i-\exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\right]\boldsymbol{z}_i-\boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma}\\&=\sum_{i=1}^n\kappa\boldsymbol{B}'_\gamma\boldsymbol{\nu}\exp(-\boldsymbol{B}_\gamma\boldsymbol{\nu})\left[y_i-\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\right]\boldsymbol{z}_i-\boldsymbol{K}_{\boldsymbol{\gamma}}\boldsymbol{\gamma}\\\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}&=\sum_{i=1}^n\kappa\exp(-\boldsymbol{B}_\gamma\boldsymbol{\nu})\left[y_i-\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\right]\boldsymbol{B}_\gamma^\top-\boldsymbol{K}_\nu\boldsymbol{\nu}\\&=\sum_{i=1}^n\kappa\exp(-\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\left[y_i-\exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\right]\boldsymbol{B}_\gamma^\top-\boldsymbol{K}_\nu\boldsymbol{\nu}\end{aligned}$$

Fisher Information of FlexGAM1 for Gamma data:

$$\begin{aligned}-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}}&=-\sum_{i=1}^n\kappa\exp(-\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\left[(y_i-\exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma})))(\Psi''(\boldsymbol{z}_i^\top\boldsymbol{\gamma})-(\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2)-(\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2\exp(\Psi(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))\right]\boldsymbol{z}_i\boldsymbol{z}_i+\boldsymbol{K}_{\boldsymbol{\gamma}}\\-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}}&=-\sum_{i=1}^n\kappa\exp(-\boldsymbol{B}_\gamma\boldsymbol{\nu})\left[(y_i-\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}))+\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\right]\boldsymbol{B}_\gamma^\top\boldsymbol{B}_\gamma^\top+\boldsymbol{K}_\nu\\-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\nu}}&=-\sum_{i=1}^n\kappa\exp(-\boldsymbol{B}_\gamma\boldsymbol{\nu})\left[(y_i-\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu}))\left(\boldsymbol{B}_\gamma'^\top-\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{B}_\gamma^\top\right)-\boldsymbol{B}'_\gamma\boldsymbol{\nu}\exp(\boldsymbol{B}_\gamma\boldsymbol{\nu})\boldsymbol{B}_\gamma^\top\right]\boldsymbol{z}_i\end{aligned}$$

Expected Fisher Information of FlexGAM1 for Gamma data:

$$\begin{aligned}F_{\boldsymbol{\gamma},\gamma}&=\mathbb{E}\left[-\frac{l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\boldsymbol{\gamma}}\right]=\sum_{i=1}^n\kappa(\Psi'(\boldsymbol{z}_i^\top\boldsymbol{\gamma}))^2\boldsymbol{z}_i\boldsymbol{z}_i+\boldsymbol{K}_{\boldsymbol{\gamma}}\\F_{\boldsymbol{\nu},\boldsymbol{\nu}}&=\mathbb{E}\left[-\frac{l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}\boldsymbol{\nu}}\right]=\sum_{i=1}^n\kappa\boldsymbol{B}_\gamma^\top\boldsymbol{B}_\gamma^\top+\boldsymbol{K}_\nu\\F_{\boldsymbol{\gamma},\boldsymbol{\nu}}&=\mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma},\boldsymbol{\nu})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\nu}}\right]=\sum_{i=1}^n\kappa\boldsymbol{B}'_\gamma\boldsymbol{\nu}\boldsymbol{B}_\gamma^\top\boldsymbol{z}_i\\F(\boldsymbol{\gamma},\boldsymbol{\nu})&=\begin{pmatrix}F_{\boldsymbol{\gamma},\gamma}&F_{\boldsymbol{\gamma},\boldsymbol{\nu}}\\F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top&F_{\boldsymbol{\nu},\boldsymbol{\nu}}\end{pmatrix}\end{aligned}$$

In the estimation procedure we fix Ψ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = (\Psi'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2$$

The "variance" κ is not necessary for the estimation part. For calculating the standard errors we use

$$\frac{1}{\kappa} = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{(\hat{\mu}_i)^2 (n-p)},$$

where p are the estimated degrees of freedom of the covariate effects.

Algorithm 9 (FlexGAM1 Gamma)

Start:

$$\hat{\gamma}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, family=Gamma(link=log))}$$

$$\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}$$

Outer (m) :

$$\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}) = \texttt{scam(}\boldsymbol{y} \sim s(\boldsymbol{\eta}^{(k-1)}\texttt{, bs="mpi"), family=Gamma(link=log))}$$

Inner (k) :

$$\boldsymbol{\mu}^{(k)} = h(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)})) = \exp(\Psi^{(m)}(\boldsymbol{\eta}^{(k-1)}))$$

$$\boldsymbol{y}^{(k)} = \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\boldsymbol{\mu}^{(k)} \Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})}$$

$$\boldsymbol{w}^{(k)} = \left(\Psi'^{(m)}(\boldsymbol{\eta}^{(k-1)})\right)^2$$

$$\hat{\boldsymbol{\gamma}}^{(k)} = \left(\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_\gamma\right)^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)}$$

$$= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \ldots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)}$$

$$\boldsymbol{\eta}_1^{(k)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}$$

A.5.2 Direct Estimation of the Response Function FlexGAM2

$$\mu_i = \mathbb{E}[y_i] = \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})$$

with

$$\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) = \sum_{l=1}^{L} B_l(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\nu_l = \boldsymbol{B}_\gamma \boldsymbol{\nu}$$

Log-likelihood of FlexGAM2 for Gamma data:

$$\begin{aligned} l(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \sum_{i=1}^{n} -\log(\Gamma(\kappa)) + \kappa \log(\kappa) - \kappa \log(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) + (\kappa - 1)\log(y_i) - \frac{\kappa y_i}{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} -\log(\Gamma(\kappa)) + \kappa \log(\kappa) - \kappa \log(\boldsymbol{B}_\gamma \boldsymbol{\nu}) + (\kappa - 1)\log(y_i) - \frac{\kappa y_i}{\boldsymbol{B}_\gamma \boldsymbol{\nu}} - \frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} - \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Score function of FlexGAM2 for Gamma data

$$\begin{aligned} \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma}} &= \sum_{i=1}^{n} \kappa \frac{\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} (y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))\boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} \\ &= \sum_{i=1}^{n} \kappa \frac{\boldsymbol{B}'_\gamma \boldsymbol{\nu}}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu})\boldsymbol{z}_i - \boldsymbol{K}_{\boldsymbol{\gamma}} \boldsymbol{\gamma} \\ \frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}} &= \sum_{i=1}^{n} \kappa \frac{1}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu})\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \\ &= \sum_{i=1}^{n} \kappa \frac{1}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} (y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))\boldsymbol{B}_\gamma^\top - \boldsymbol{K}_\nu \boldsymbol{\nu} \end{aligned}$$

Fisher Information of FlexGAM2 for Gamma data:

$$\begin{aligned} -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}} &= -\sum_{i=1}^{n} \kappa \left[\frac{\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})\hat{h}''(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) - 2(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^3} (y_i - \hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) - \frac{(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} \right] \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} \kappa \left[(-2)\frac{y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^3} - \frac{1}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} \right] \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ -\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} &= -\sum_{i=1}^{n} \kappa \left(\frac{\boldsymbol{B}_\gamma \boldsymbol{\nu} \boldsymbol{B}_\gamma'^\top - 2\boldsymbol{B}'_\gamma \boldsymbol{\nu} \boldsymbol{B}_\gamma^\top}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^3} (y_i - \boldsymbol{B}_\gamma \boldsymbol{\nu}) - \frac{\boldsymbol{B}'_\gamma \boldsymbol{\nu}}{(\boldsymbol{B}_\gamma \boldsymbol{\nu})^2} \boldsymbol{B}_\gamma^\top \right) \boldsymbol{z}_i \end{aligned}$$

Expected Fisher Information of FlexGAM2 for Gamma data:

$$\begin{aligned} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \boldsymbol{\gamma}} \right] &= \sum_{i=1}^{n} \kappa \frac{(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} \boldsymbol{z}_i \boldsymbol{z}_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \\ F_{\boldsymbol{\nu},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \boldsymbol{\nu}} \right] &= \sum_{i=1}^{n} \kappa \frac{1}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} \boldsymbol{B}_\gamma^\top \boldsymbol{B}_\gamma^\top + \boldsymbol{K}_\nu \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}} = \mathbb{E}\left[-\frac{\partial l(\boldsymbol{\gamma}, \boldsymbol{\nu})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\nu}} \right] &= \sum_{i=1}^{n} \kappa \frac{\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2} \boldsymbol{B}_\gamma^\top \boldsymbol{z}_i \\ F(\boldsymbol{\gamma}, \boldsymbol{\nu}) &= \begin{pmatrix} F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} & F_{\boldsymbol{\gamma},\boldsymbol{\nu}} \\ F_{\boldsymbol{\gamma},\boldsymbol{\nu}}^\top & F_{\boldsymbol{\nu},\boldsymbol{\nu}} \end{pmatrix} \end{aligned}$$

In the estimation procedure we fix $\hat{h}$ to estimate $\boldsymbol{\gamma}$, thus we focus on $F_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ to get the working weights:

$$\boldsymbol{F}(\boldsymbol{\gamma}) = F_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i w_i + \boldsymbol{K}_{\boldsymbol{\gamma}} \qquad \text{with } w_i = \frac{(\hat{h}'(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}{(\hat{h}(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))^2}$$

The "variance" κ is not necessary for the estimation part. For calculating the standard errors we use

$$\frac{1}{\kappa} = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{(\hat{\mu}_i)^2 (n-p)},$$

where p are the estimated degrees of freedom of the covariate effects.

Algorithm 10 (FlexGAM2 Gamma)

Start:

$$\hat{\boldsymbol{\gamma}}^{(0)} = \texttt{gam(}\boldsymbol{y} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots\texttt{, family=Gamma(link=log))}$$

$$\boldsymbol{\eta}_1^{(0)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(0)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(0)} = \frac{\boldsymbol{\eta}_1^{(0)} - \text{mean}(\boldsymbol{\eta}_1^{(0)})}{\text{sd}(\boldsymbol{\eta}_1^{(0)})}$$

Outer (m) :

$$\hat{\boldsymbol{\nu}}^{(m)} = \text{argmin} \left(\sum_{i=1}^{n} \left(y_i - \sum_{l=1}^{L} B_l(\eta_i^{(k-1)}) \nu_l \right)^2 + \boldsymbol{\nu}^\top \boldsymbol{K}_\nu \boldsymbol{\nu} \right) \Rightarrow \hat{h}^{(m)}$$

Inner (k) :

$$\boldsymbol{\mu}^{(k)} = \hat{h}^{(m)}(\boldsymbol{\eta}^{(k-1)})$$

$$\boldsymbol{y}^{(k)} = \boldsymbol{\eta}^{(k-1)} + \frac{\boldsymbol{y} - \boldsymbol{\mu}^{(k)}}{\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})}$$

$$\boldsymbol{w}^{(k)} = \left(\frac{\hat{h}'^{(m)}(\boldsymbol{\eta}^{(k-1)})}{\hat{h}^{(m)}(\boldsymbol{\eta}^{(k-1)})} \right)^2$$

$$\hat{\boldsymbol{\gamma}}^{(k)} = \left(\boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{Z} + \boldsymbol{K}_\gamma \right)^{-1} \boldsymbol{Z}^\top \boldsymbol{W}^{(k)} \boldsymbol{y}^{(k)}$$

$$= \texttt{gam(}\boldsymbol{y}^{(k)} \sim \boldsymbol{x}_1 + \boldsymbol{x}_2 + s(\boldsymbol{x}_3) + s(\boldsymbol{x}_4) + \dots\texttt{, weight = }\boldsymbol{w}^{(k)}\texttt{, family=gaussian)}$$

$$\boldsymbol{\eta}_1^{(k)} = \boldsymbol{Z}\hat{\boldsymbol{\gamma}}^{(k)} \qquad \Rightarrow \qquad \boldsymbol{\eta}^{(k)} = \frac{\boldsymbol{\eta}_1^{(k)} - \text{mean}(\boldsymbol{\eta}_1^{(k)})}{\text{sd}(\boldsymbol{\eta}_1^{(k)})}$$

However in the Gamma case the constraints for the estimation of $\hat{h}$ are that the spline has to be strictly monotone increasing and positive. There is no upper bound for the fitted value and the spline. So the coefficients $\boldsymbol{\nu}$ have to fulfill:

$$0 < \nu_l < \nu_{l+1}$$

B Simulation Study

In this section we provide all results of the simulation study. The ordering is that we start with the results for the models with linear predictor and end with the results of the models with smooth predictors. In each subsection we first present the goodness-of-fit criteria, then we display the estimated response functions, continue with the estimated covariate effects and conclude with the estimated ROC curves. The effects are scaled with $\frac{\eta-\bar{\eta}}{\mathrm{sd}(\eta)}$ to be comparable.

B.1 Linear Predictor

B.1.1 Goodness-of-fit

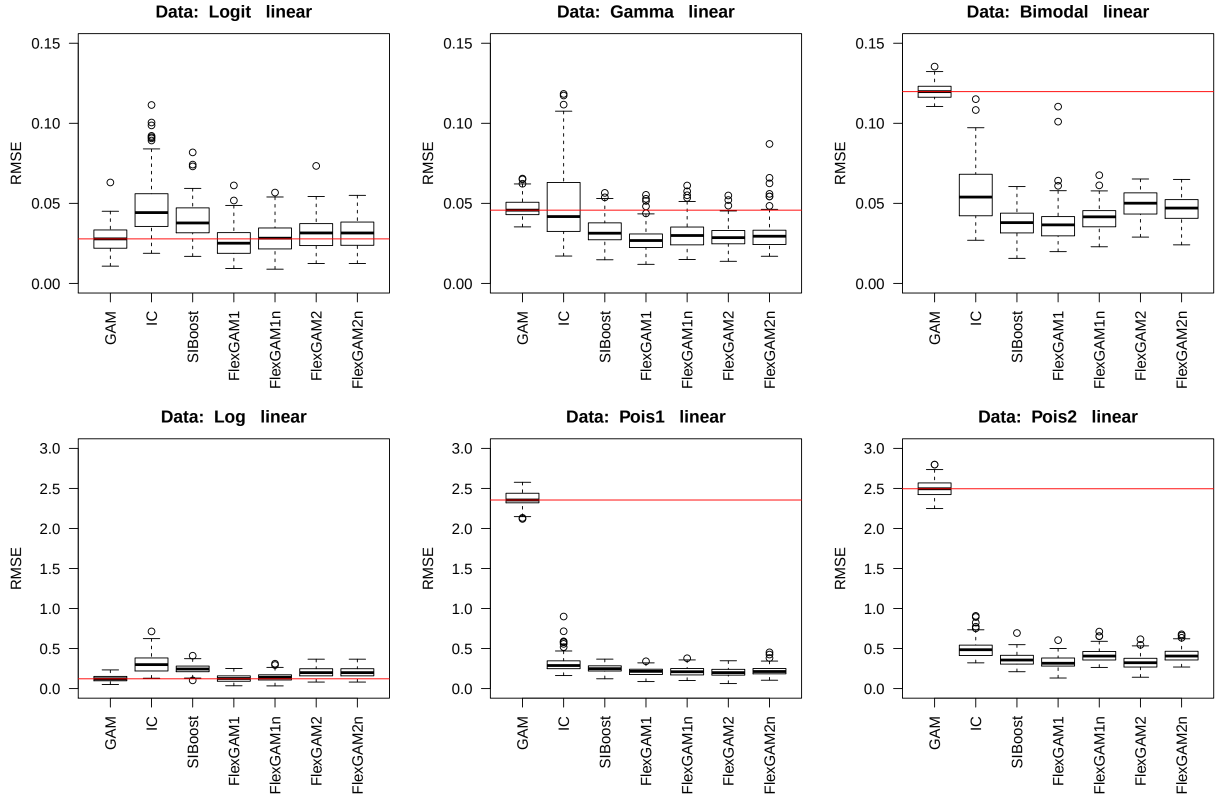

Figure B.1: RMSE of the models with linear predictor

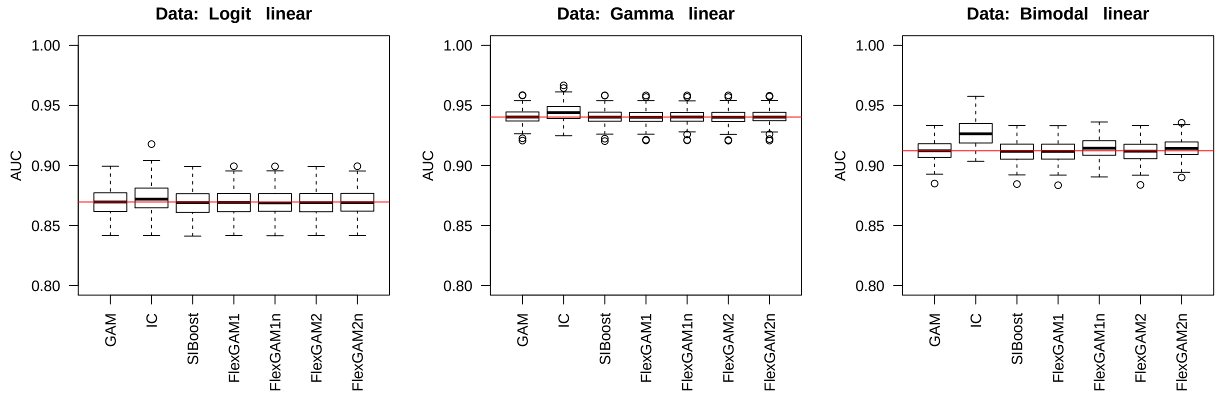

Figure B.2: AUC of the binomial models with linear predictor

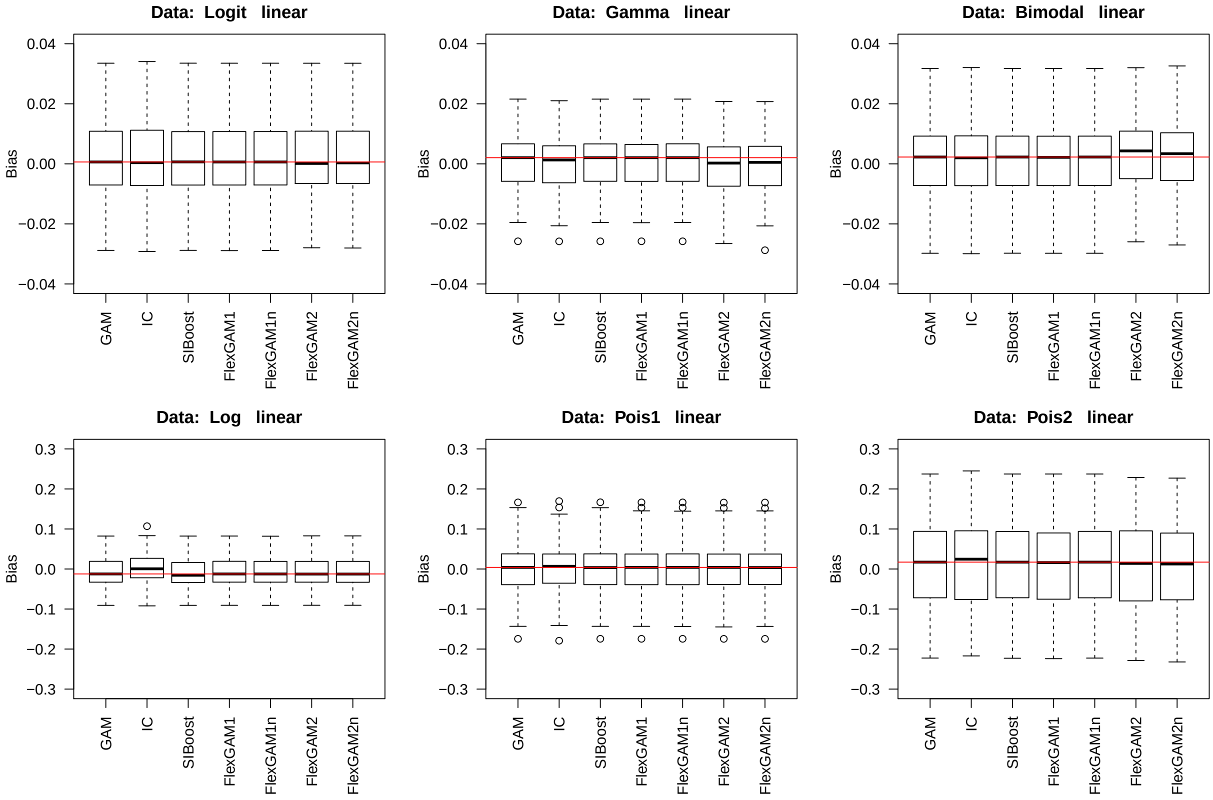

Figure B.3: Bias of the models with linear predictor

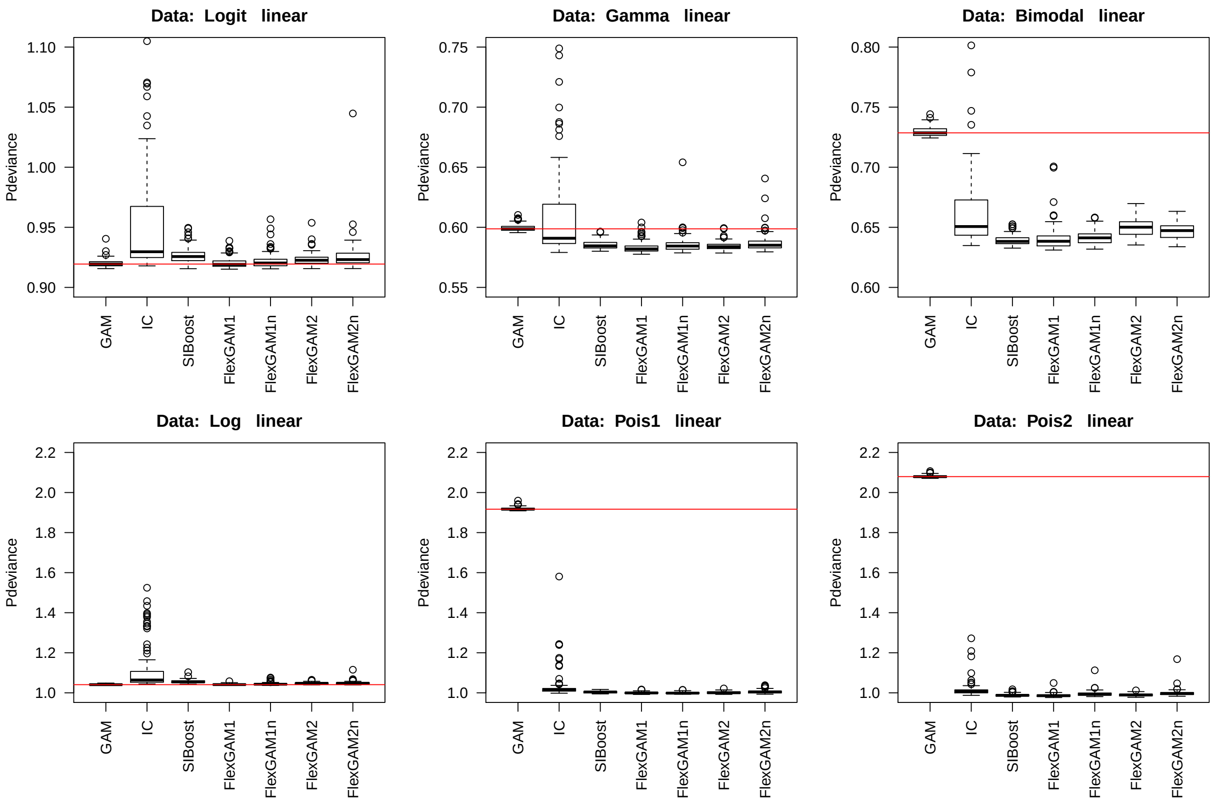

Figure B.4: Predictive Deviance of the models with linear predictor (scaled by 10000)

B.1.2 Estimated Response Functions

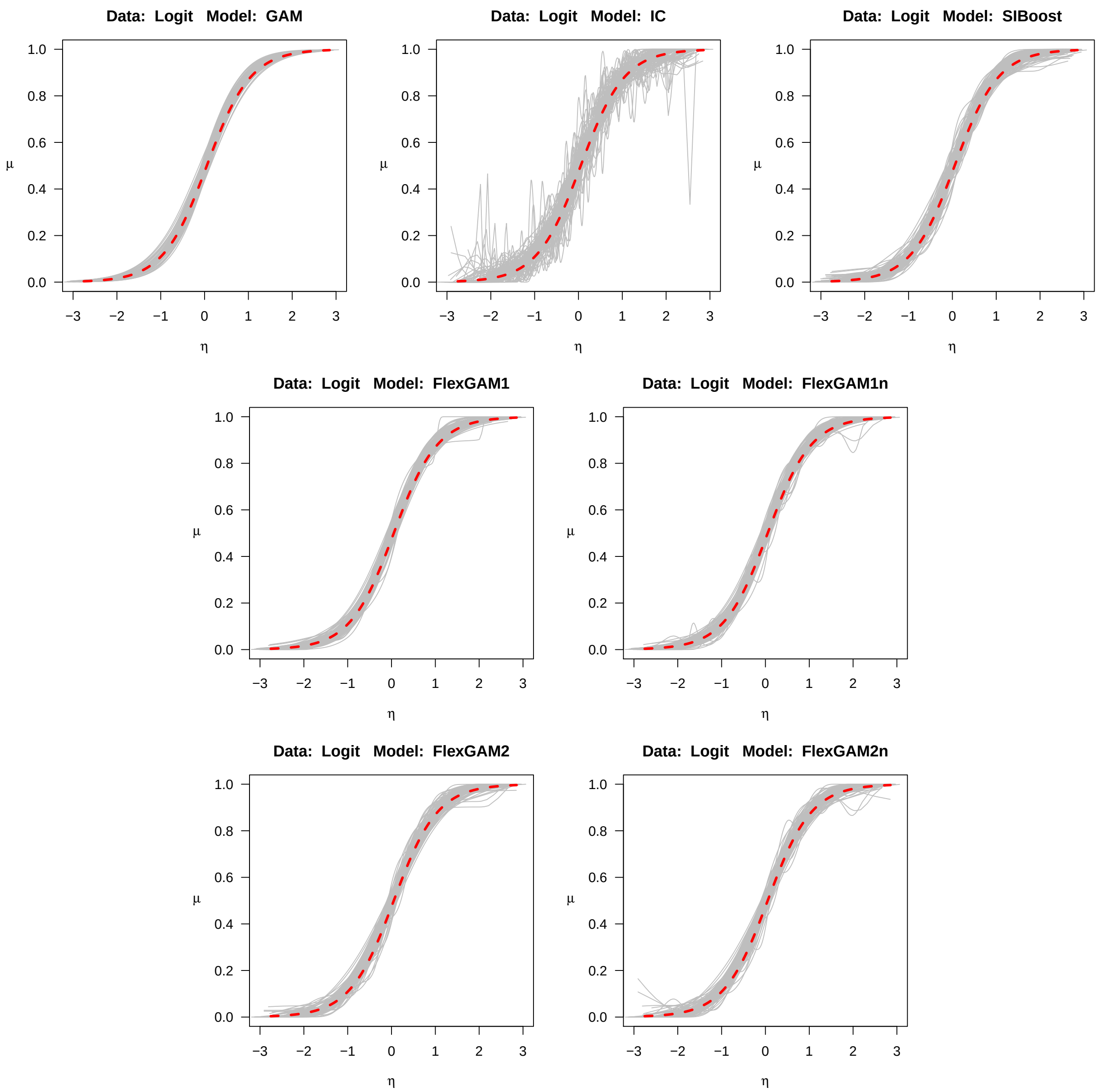

Figure B.5: Estimated response functions for *Logit* data and of the models with linear predictor

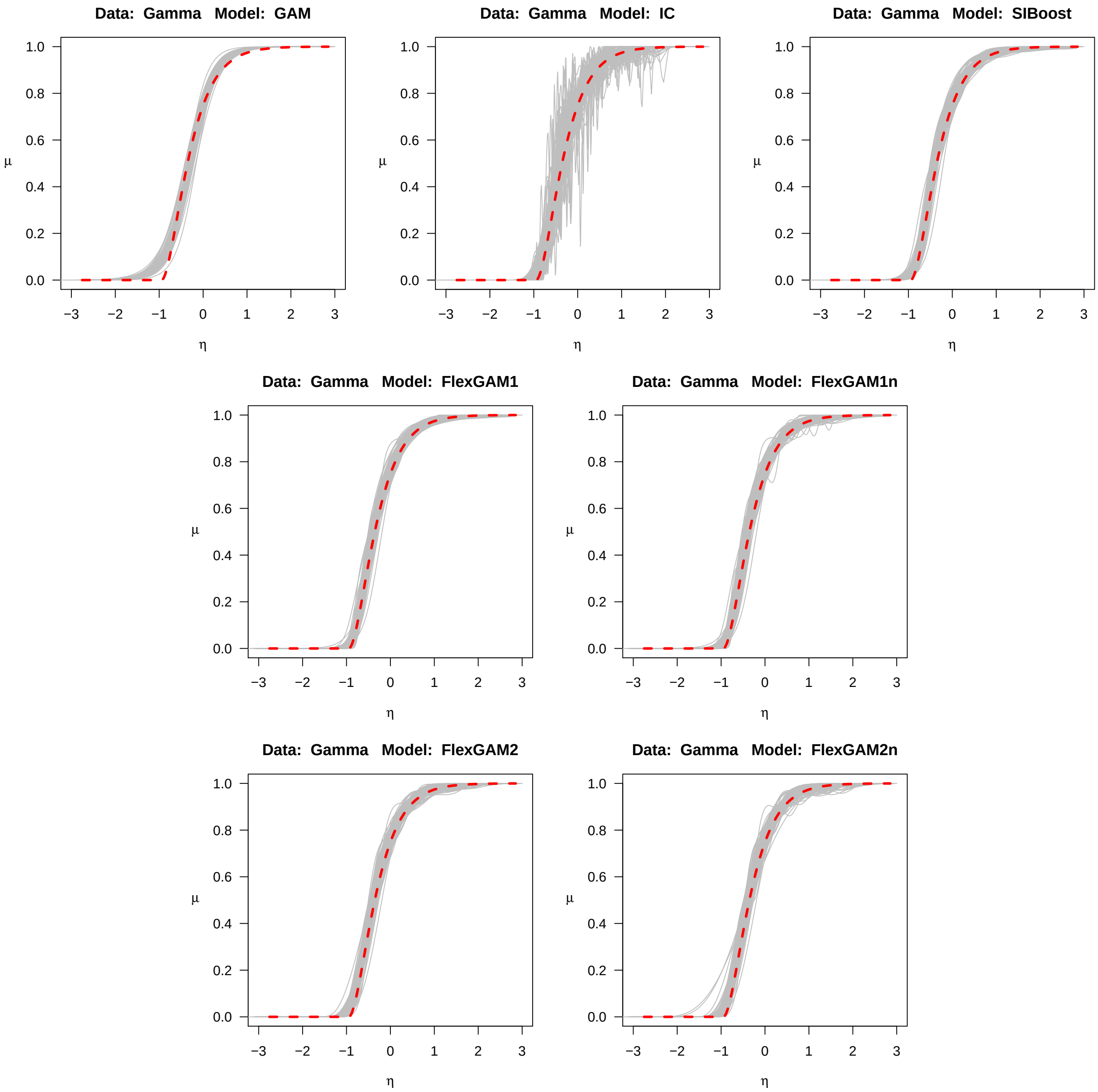

Figure B.6: Estimated response functions for *Gamma* data and of the models with linear predictor

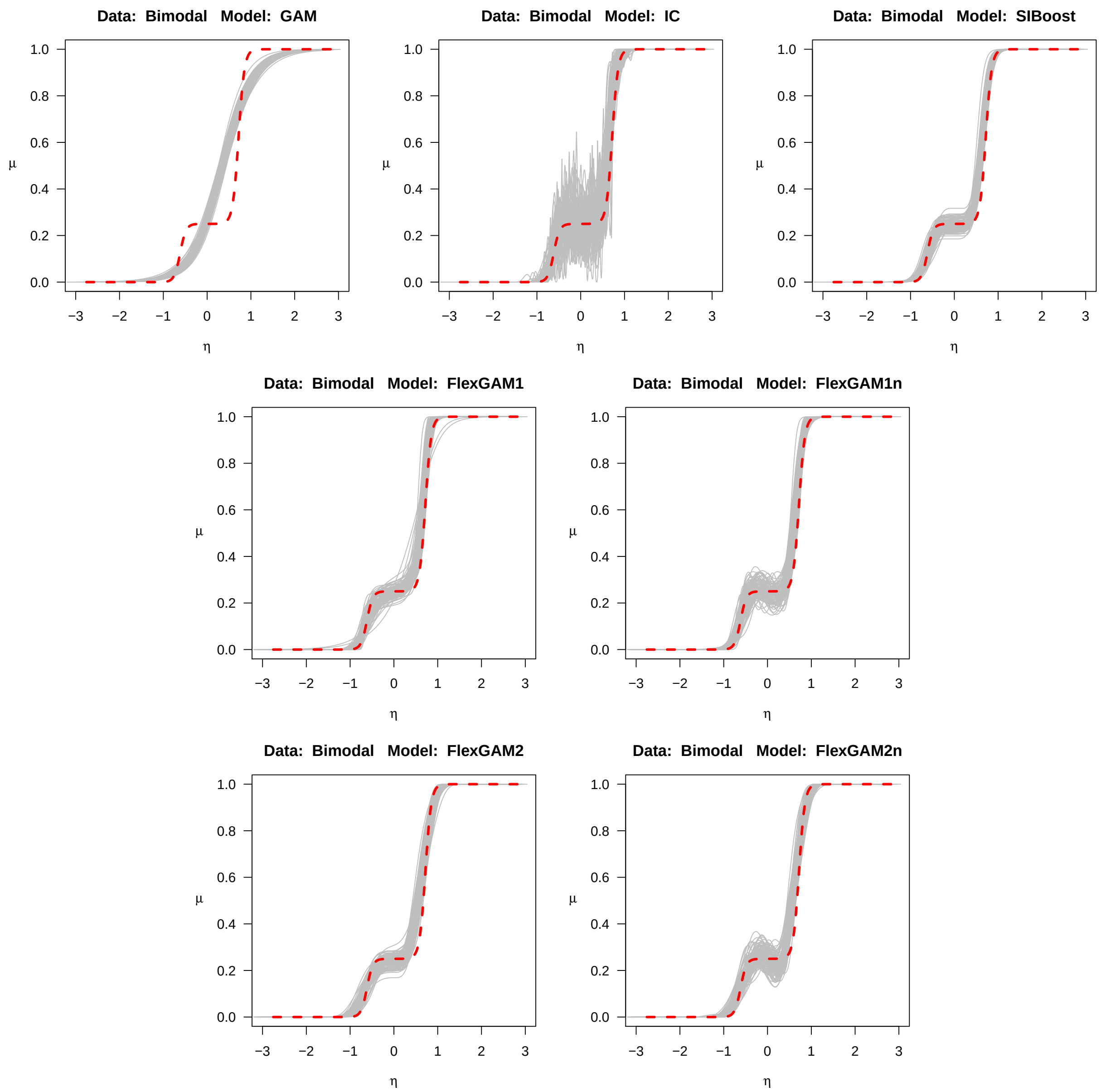

Figure B.7: Estimated response functions for *Bimodal* data and of the models with linear predictor

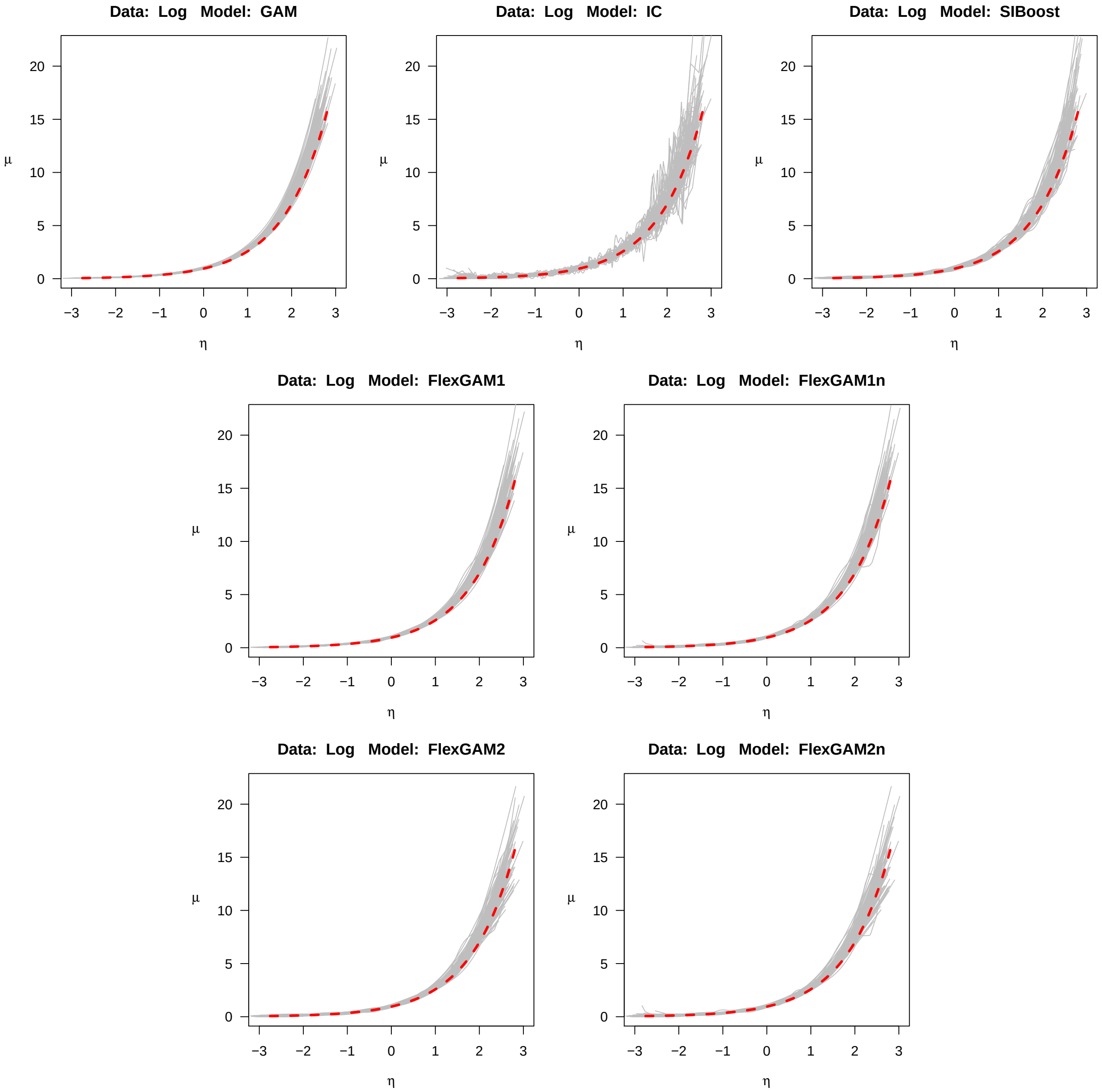

Figure B.8: Estimated response functions for *Log* data and of the models with linear predictor

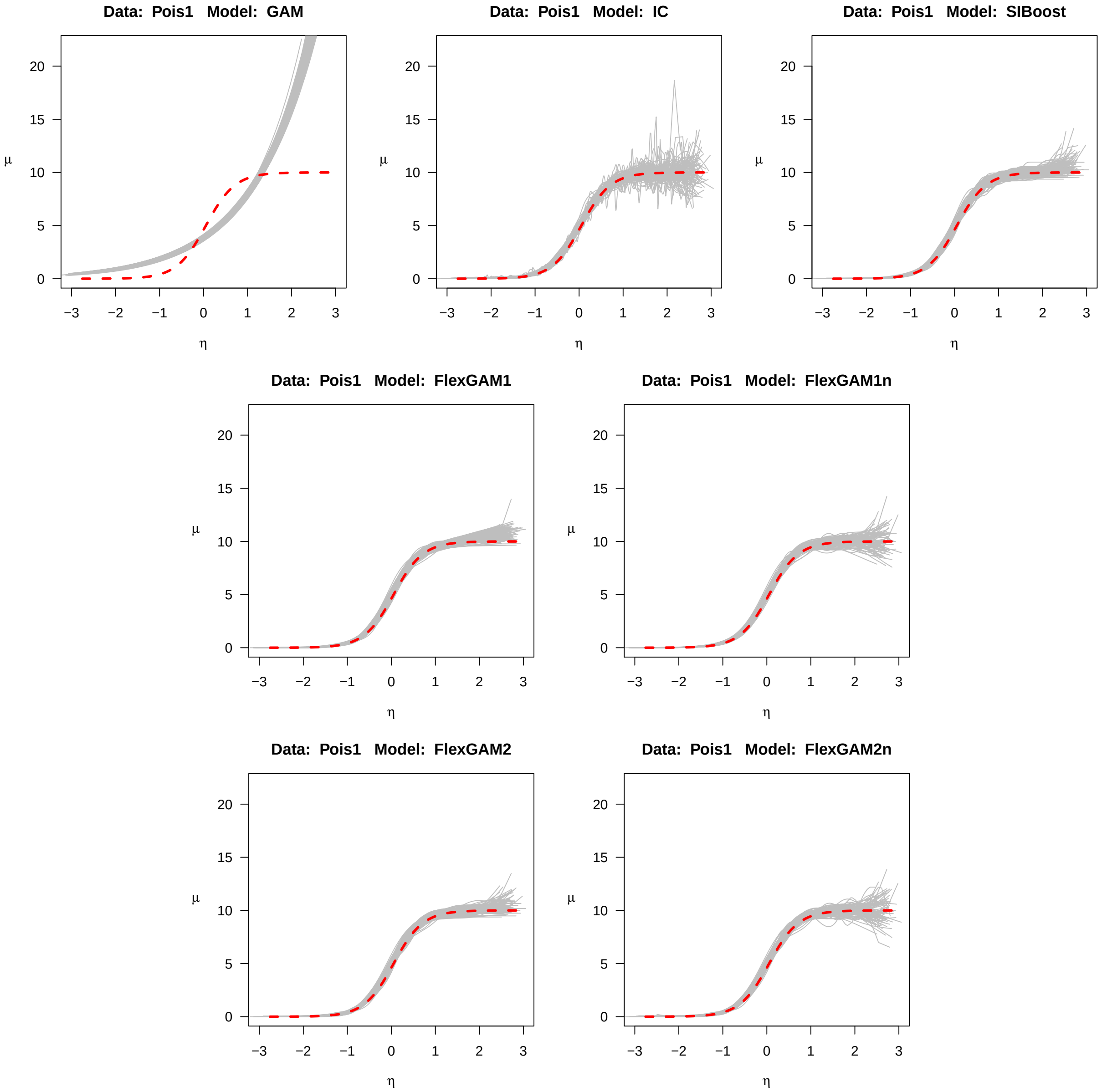

Figure B.9: Estimated response functions for *Pois1* data and of the models with linear predictor

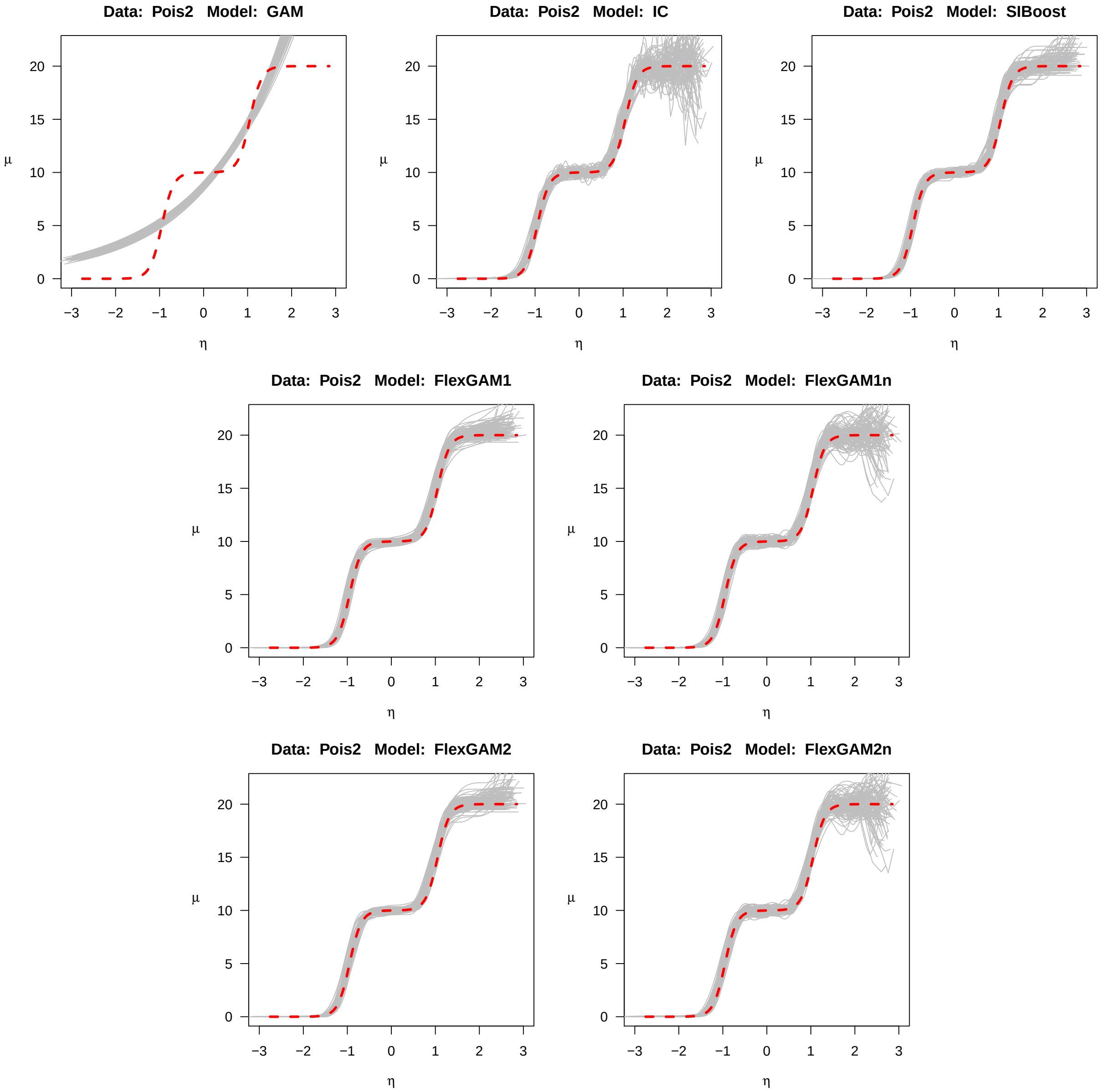

Figure B.10: Estimated response functions for *Pois2* data and of the models with linear predictor

B.1.3 Estimated Covariate Effects

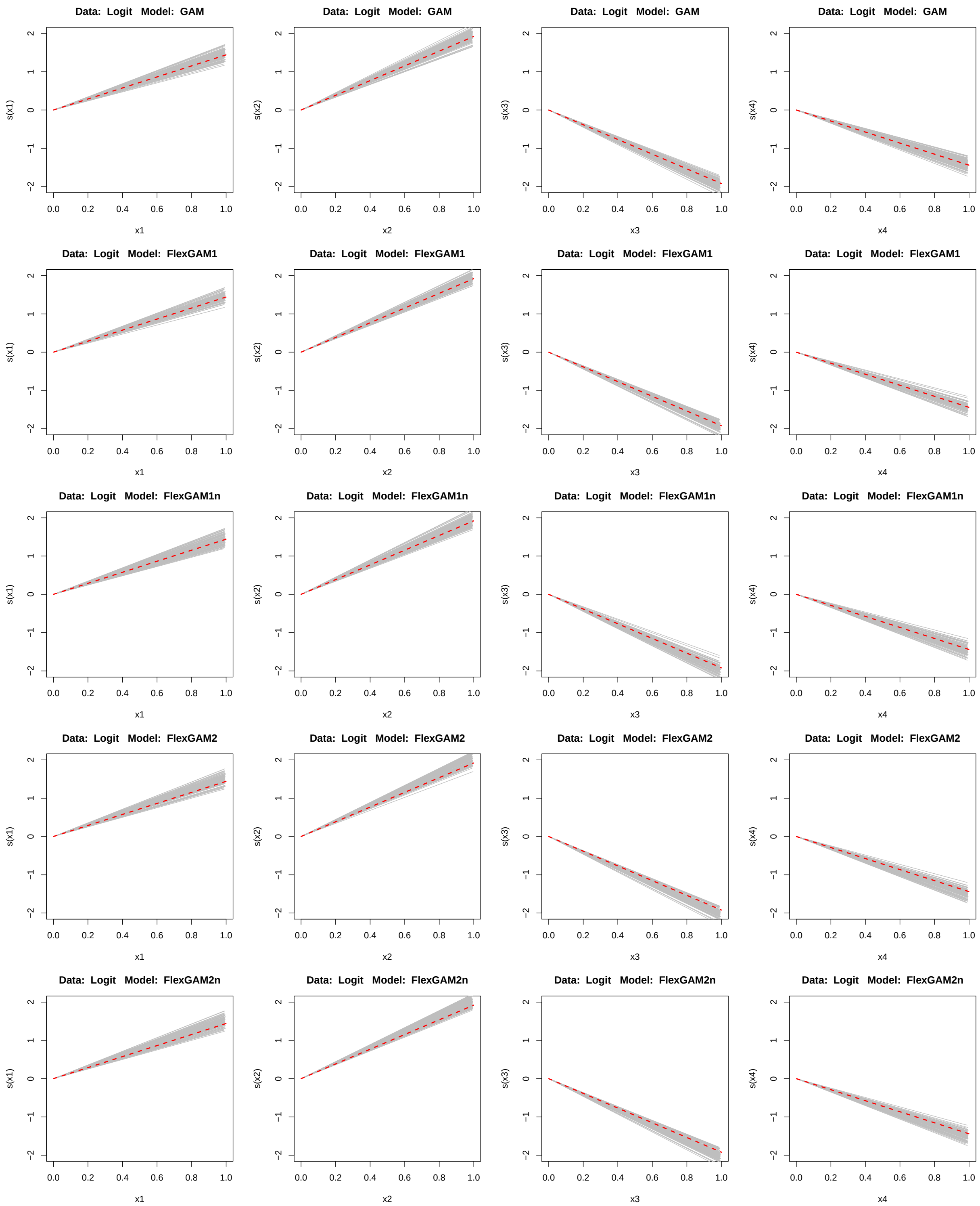

Figure B.11: Estimated scaled covariate effects for *Logit* data and of the models with linear predictor

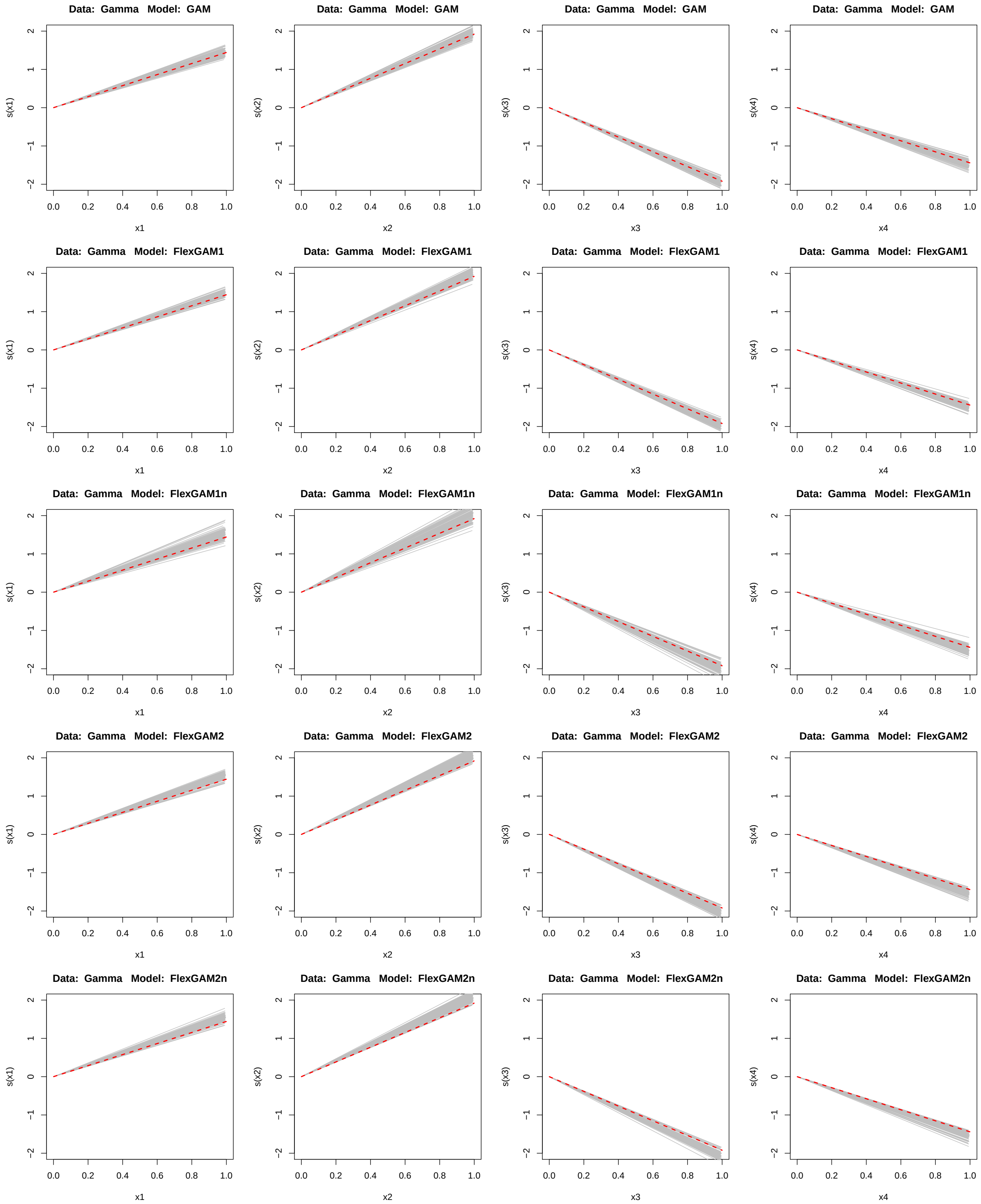

Figure B.12: Estimated scaled covariate effects for *Gamma* data and of the models with linear predictor