



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Czado, Raftery:

## Choosing the Link Function and Accounting for Link Uncertainty in Generalized Linear Models using Bayes Factors

Sonderforschungsbereich 386, Paper 262 (2001)

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

Projektpartner



# Choosing the Link Function and Accounting for Link Uncertainty in Generalized Linear Models using Bayes Factors

**Claudia Czado\***

Zentrum Mathematik, Technische Universität München,  
Boltzmannstr.3 , D-85747 Garching bei München, Germany

**and**

**Adrian E. Raftery†**

Department of Statistics, University of Washington,  
Box 354322, Seattle, WA 98195-4322, USA

June 17, 2003

ABSTRACT. One important component of model selection using generalized linear models (GLM) is the choice of a link function. We propose using approximate Bayes factors to assess the improvement in fit over a GLM with canonical link when a parametric link family is used. The approximate Bayes factors are calculated using the Laplace approximations given in Raftery (1996), together with a reference set of prior distributions. This methodology can be used to differentiate between different parametric link families, as well as allowing one to jointly select the link family and the independent variables.

---

\* *email:* cczado@ma.tum.de

† *email:* raftery@stat.washington.edu

This involves comparing nonnested models and so standard significance tests cannot be used. The approach also accounts explicitly for uncertainty about the link function. The methods are illustrated using parametric link families studied in Czado (1997) for two data sets involving binomial responses.

KEY WORDS: Bayes factors; link function; GLM, model selection; reference prior.

## 1. Introduction

To find an appropriate generalized linear model (GLM) for regression data involves choosing the independent variables, the link function and the variance function (McCullagh and Nelder (1989)). Typically, many different models are compared using individual significance tests based on the asymptotic distribution of the deviance. As pointed out by Gelfand and Dey (1994) and Raftery (1996), this strategy cannot be used for comparing nonnested models. In addition, model uncertainty is usually ignored, as are power considerations. A Bayesian approach can avoid these difficulties and to implement this, Raftery (1996) developed approximate Bayes factors for GLM's based on the Laplace method for integrals. These approximations require only the maximum likelihood estimate (MLE), the deviance and the observed or expected Fisher information. Kass and Raftery (1995) and Han and Carlin (2001) review Bayes factors and discuss different ways to calculate Bayes factors.

In this paper, we extend the approach taken by Raftery (1996) to calculate approximate Bayes factors for GLM's with a parametric link function. Even though GLM's with canonical links (for definition see McCullagh and Nelder (1989)), such as the logit link in binomial regression, guarantee maximum

information and a simple interpretation of the regression parameters, they do not always provide the best fit available to a given data set. Link misspecification can lead to substantial bias in the regression parameters and the mean response estimates (see Czado and Santner (1992) for binomial responses). One common approach to guard against link misspecification in generalized linear models is to embed the canonical link in a wide parametric class of links  $\mathfrak{S} = \{F(\cdot, \psi), \psi \in \Psi\}$ , which includes the canonical link as a special case when  $\psi = \psi_0$ . Many such parametric link classes for binary regression data have been proposed in the literature. Montfort and Otten (1976), Copenhaver and Mielke (1977), Aranda-Ordaz (1981), Guerrero and Johnson (1982), Morgan (1983) and Whittmore (1983) proposed one-parameter families, while Prentice (1976), Pregibon (1980), Stukel (1988) and Czado (1992) considered two-parameter families. Link functions for the non-binary case were studied by Pregibon (1980) and Czado (1992, 1997).

With the multitude of link families to choose from, the Bayes factor approach is able to compare different link families, regardless of whether they are nested or nonnested. We will illustrate this ability by using the two-parameter link family suggested by Czado (1997) in several data sets. In addition, we are able to choose the link family and the set of independent variables jointly.

In Section 2 we define and discuss GLM's with parametric links, while in Section 3 the calculation of approximate Bayes factors including the choice of priors will be discussed. Applications will be given in Section 4 and Section 5 will provide a summary and discussion of the method presented.

## 2. Generalized Linear Models with Parametric Links

The following model for regression data with response  $Y_i$  and independent variables  $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})$  for  $i = 1, \dots, n$  will be used:

1. **Random Component:**  $\{Y_i, 1 \leq i \leq n\}$  are independent and have a density of the form

$$f_{y_i}(y_i, \theta_i, \phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right], \quad (2.1)$$

for some specified functions  $a(\cdot), b(\cdot)$  and  $c(\cdot)$ . The scale parameter  $\phi$  is allowed to be known or unknown.

2. **Systematic Component:** The linear predictors  $\eta_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$  for  $1 \leq i \leq n$  influence the response  $Y_i$ . Here  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$  are unknown regression parameters.
3. **Parametric Link Component:** The linear predictors  $\eta_i(\boldsymbol{\beta})$  are related to the mean  $\mu_i$  of  $Y_i$  by  $\mu_i = F(\eta_i(\boldsymbol{\beta}), \psi)$  for some  $F(\cdot, \psi)$  in  $\mathfrak{S} = \{F(\cdot, \psi) : \psi \in \Psi\}$ .

We will restrict attention to link families  $\mathfrak{S}$  that contain only strictly monotone continuous functions  $F(\cdot, \psi)$ . Note that in conventional GLM notation the link  $g$  is equal to the inverse of  $F$ . An unknown scale parameter  $\phi$  in (2.1) is typically estimated by an appropriate moment estimator involving the Pearson  $\chi^2$  Statistic (McCullagh and Nelder (1989)). For a fixed link parameter  $\psi$  we remain in the class of GLM's, while this is no longer true if the link parameter  $\psi$  and the regression parameter  $\boldsymbol{\beta}$  are jointly estimated by the data. Czado and Munk (2000) show that the joint MLE  $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}, \hat{\psi})$  of  $\boldsymbol{\delta} = (\boldsymbol{\beta}, \psi)$  is strongly consistent and efficient under regularity conditions.

**Table 1**  
*Link Families for GLM's*

Error Distribution	Parameter Restriction	Canonical Link	Link Family $\mathfrak{S} = \{F(\cdot, \psi) : \psi \in \Psi\}$
Normal	$\mu$ real	$F(\eta) = \eta$	$F(\eta, \psi) = h(\eta, \psi)$
Binomial	$\mu \in (0, 1)$	$F(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}$	$F(\eta, \psi) = \frac{\exp(h(\eta, \psi))}{1 + \exp(h(\eta, \psi))}$
Poisson	$\mu > 0$	$F(\eta) = \exp(\eta)$	$F(\eta, \psi) = \exp(h(\eta, \psi))$
Gamma	$\mu > 0$	$F(\eta) = \eta^{-1}$	$F(\eta, \psi) = [\exp(h(\eta, \psi))]^{-1}$
Inv. Gaussian	$\mu > 0$	$F(\eta) = \eta^{-.5}$	$F(\eta, \psi) = [\exp(h(\eta, \psi))]^{-.5}$

We will illustrate our approach by using the link families suggested by Czado (1997). They allow separate modifications of the left and/or right tail of the link function and exhibit low variance inflation (Taylor (1988), Taylor et al. (1996)) for the regression parameters when the link is estimated from the data. This is due to the fact that the parametrization is locally orthogonal (see Cox and Reid (1987)). In addition, they are location and scale invariant (see Czado (1997)). For GLM's with parametric links they are defined as in Table 1. In Table 1,  $h(\eta, \psi)$  is one of the following functions:

$$\text{Both tails: } h_b(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) = \begin{cases} +\frac{(\eta+1)^{\psi_1}-1}{\psi_1} & \text{if } \eta \geq 0 \\ -\frac{(-\eta+1)^{\psi_2}-1}{\psi_2} & \text{otherwise} \end{cases} \quad (2.2)$$

$$\text{Right tail: } h_r(\eta, \psi_1) = \begin{cases} +\frac{(\eta+1)^{\psi_1}-1}{\psi_1} & \text{if } \eta \geq 0 \\ \eta & \text{otherwise} \end{cases} \quad (2.3)$$

$$\text{Left tail: } h_l(\eta, \psi_2) = \begin{cases} \eta & \text{if } \eta \geq 0 \\ -\frac{(-\eta+1)^{\psi_2}-1}{\psi_2} & \text{otherwise} \end{cases} \quad (2.4)$$

The parameter restriction for the mean response makes a right tail modification for the Poisson and a left tail modification for the Gamma and inverse Gaussian cases the only sensible modifications to be considered. In all other cases all modifications of the link function are allowed. In particular, (2.4)

is a special case of (2.2) with  $\psi_1 = 1$ , and (2.3) is a special case of (2.2) with  $\psi_2 = 1$ . As  $\psi_1$  increases, the right tail of  $G(\cdot, \boldsymbol{\psi})$  becomes lighter, while an increasing  $\psi_2$  makes the left tail of  $G(\cdot, \boldsymbol{\psi})$  lighter. The specification (2.3) is asymmetric if  $\psi_1 \neq 1$ , while the specification (2.4) is asymmetric if  $(\psi_2 \neq 1)$ . The both tails specification (2.2) is asymmetric if  $\psi_1 \neq \psi_2$ . Further, for  $\psi_1 < 0$  and  $\psi_2 < 0$

$$\begin{aligned} \lim_{\eta \rightarrow \infty} h_r(\eta, \psi_1) &= \frac{1}{|\psi_1|} \text{ and } \lim_{\eta \rightarrow -\infty} h_l(\eta, \psi_2) = -\frac{1}{|\psi_2|}, \\ \lim_{\eta \rightarrow \infty} h_b(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) &= \frac{1}{|\psi_1|} \text{ and } \lim_{\eta \rightarrow -\infty} h_b(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) = -\frac{1}{|\psi_2|}. \end{aligned}$$

This, together with monotonicity of  $h_r, h_l$  and  $h_b$ , imply restrictions on the range of allowable means  $\mu_i = E(Y_i)$  if  $\psi_1 < 0$  or  $\psi_2 < 0$ . In particular, for binomial links, we have restrictions on the allowable success probabilities  $p_i$  given by

$$p_i \leq \frac{\exp\{\frac{1}{|\psi_1|}\}}{1 + \exp\{\frac{1}{|\psi_1|}\}} \text{ if } \psi_1 < 0 \quad (2.5)$$

$$p_i \leq \frac{\exp\{-\frac{1}{|\psi_2|}\}}{1 + \exp\{-\frac{1}{|\psi_2|}\}} \text{ if } \psi_2 < 0. \quad (2.6)$$

For example if we want an allowable range of success probabilities between .1 and .9, this implies that  $\psi_1 \geq -.46$  and  $\psi_2 \geq -.46$  if negative link values are allowed, by inversion of (2.5) and (2.6). There are no restrictions on the success probabilities when  $\psi_1 \geq 0$  and  $\psi_2 \geq 0$ .

### 3. Approximate Bayes Factors for GLM's with Parametric Link

We are interested in assessing the evidence for a GLM with a noncanonical link as against the same GLM with a canonical link using Bayes factors. For this, we denote by  $M_\psi$  a GLM with a fixed link parameter  $\psi$  for a given set of

independent variables, while  $M_c$  denotes the same GLM using the canonical link. We denote the regression parameter corresponding to model  $M_\psi$  by  $\boldsymbol{\beta}_\psi$  to indicate that the regression parameters are on different scales for different  $\psi$ 's. We are interested in the Bayes factor for model  $M_\psi$  against model  $M_c$  given the data  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , which is defined as the ratio of posterior to prior odds, namely

$$B_\psi := \frac{pr(\mathbf{Y}|M_\psi)}{pr(\mathbf{Y}|M_c)}, \quad (3.1)$$

the ratio of the integrated likelihoods. In equation (3.1),

$$pr(\mathbf{Y}|M_\psi) = \int pr(\mathbf{Y}|M_\psi, \boldsymbol{\beta}_\psi)p(\boldsymbol{\beta}_\psi|M_\psi)d\boldsymbol{\beta}_\psi, \quad (3.2)$$

where  $\boldsymbol{\beta}_\psi$  is the corresponding regression parameter in Model  $M_\psi$  and  $p(\boldsymbol{\beta}_\psi|M_\psi)$  is its prior density in model  $M_\psi$ . Note that  $M_c$  corresponds to  $M_\psi$  with  $\psi = 1$ .

The Bayes factor is a summary of the evidence for  $M_\psi$  against  $M_c$  provided by the data. Sometimes it is useful to consider  $2 \log B_\psi$ , which is on the same scale as the familiar deviance and likelihood ratio test statistics. We use the rounded scale given in Table 1 of Raftery (1996) for interpreting  $B_\psi$  or  $2 \log B_\psi$ .

This approach allows us to compare different parametric link families as follows. Let  $M_\theta$  denote a GLM using a link family indexed by the link parameter  $\theta$  and construct  $B_\theta$  in a similar fashion as  $B_\psi$ . The quantity  $\frac{B_\psi}{B_\theta}$  then provides a summary of the evidence for model  $M_\psi$  against model  $M_\theta$  given the data and the same set of independent variables. In a similar way we can construct comparisons of models with different sets of independent variables and link parameters.

For the link families given in Table 1 it is also of interest to assess whether



a right tail, left tail or a both tail modification is needed. For this we can compare  $B_{\psi_1}(B_{\psi_2})$  and  $B_{\boldsymbol{\psi}=(\psi_1,\psi_2)}$  for individual link parameter values or construct overall Bayes factors for each tail modification, given by

$$\text{Both Tails : } B_b = \int B_{\boldsymbol{\psi}=(\psi_1,\psi_2)} pr(\boldsymbol{\psi}|M_{\boldsymbol{\psi}=(\psi_1,\psi_2)}) d\boldsymbol{\psi} \quad (3.3)$$

$$\text{Right Tail : } B_r = \int B_{\psi_1} pr(\psi_1|M_{\psi_1}) d\psi_1 \quad (3.4)$$

$$\text{Left Tail : } B_l = \int B_{\psi_2} pr(\psi_2|M_{\psi_2}) d\psi_2, \quad (3.5)$$

where  $pr(\boldsymbol{\psi}|M_{\boldsymbol{\psi}=(\psi_1,\psi_2)})$ ,  $pr(\psi_1|M_{\psi_1})$  and  $pr(\psi_2|M_{\psi_2})$  denote the corresponding prior densities for  $\boldsymbol{\psi}$ ,  $\psi_1$  and  $\psi_2$ , respectively. If the link parameter values are not chosen in advance, but instead are estimated,  $B_{\psi_1}$ ,  $B_{\psi_2}$  and  $B_{\boldsymbol{\psi}}$  will tend to overstate the evidence for a modification. The overall Bayes factors  $B_r$ ,  $B_l$  and  $B_b$  are preferable in this case, because they take into account the fact that the link parameters are unknown and thus take link uncertainty into account. For example, the ratio  $\frac{B_b}{B_r}$  will compare a both tails modification to a right tail one. In a similar fashion we can assess the evidence for one link family against another one given the same or different set of independent variables.

To complete the specification of these overall Bayes factors, we have to select prior distributions for the regression parameters given a model with a specified link parameter, as well as the prior distribution to be used for the link parameter.

For the prior distribution of the regression parameters  $\boldsymbol{\beta}_\psi$  in the model  $M_\psi$  we use the reference proper prior distributions suggested by Raftery (1996) for GLM's, since for fixed values of the link parameter  $\psi$  we remain in the class of ordinary GLM's. These prior distributions assume little prior

information. They are based on adjusted dependent variables to mimic the behavior for ordinary linear regression models. For a  $(p + 1)$ -dimensional  $\beta_\psi$  including an intercept, we use the prior

$$\beta_\psi | M_\psi \sim N_{p+1}(\mathbf{v}_\psi, Q_\psi U Q_\psi'), \quad (3.6)$$

where  $N_p(\boldsymbol{\mu}, \Sigma)$  denotes a  $p$ -dimensional normal distribution with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ . To specify the quantities in (3.6), the adjusted dependent variable  $z_i^\psi = g_\psi(\hat{\mu}_i^\psi) + (y_i - \hat{\mu}_i^\psi)g'_\psi(\hat{\mu}_i^\psi)$  with weights  $w_i^\psi$  (McCullagh and Nelder (1989), p.40) has to be considered. Here  $\hat{\mu}_i^\psi$  denotes the MLE of the  $i$ th mean response in the GLM with link parameter  $\psi$ , and  $g_\psi(\cdot)$  is the inverse of  $F(\cdot, \psi)$ . Define the weighted summary statistics:

$$\bar{z}_\psi = \frac{\sum_{i=1}^n w_i^\psi z_i^\psi}{\sum_{i=1}^n w_i^\psi} \text{ and } s_0^\psi = \sqrt{\frac{\sum_{i=1}^n w_i^\psi (z_i^\psi - \bar{z}_\psi)^2}{\sum_{i=1}^n w_i^\psi}}, \quad (3.7)$$

$$\bar{x}_j^\psi = \frac{\sum_{i=1}^n w_i^\psi x_{ij}}{\sum_{i=1}^n w_i^\psi} \text{ and } s_j^\psi = \sqrt{\frac{\sum_{i=1}^n w_i^\psi (x_{ij} - \bar{x}_j^\psi)^2}{\sum_{i=1}^n w_i^\psi}}, \quad j = 1 \cdots, p. \quad (3.8)$$

Then the prior mean is specified as  $\mathbf{v}'_\psi = (\bar{z}_\psi, 0, \dots, 0)'$ ,  $U$  denotes a diagonal matrix with diagonal entries given by  $(1, \sigma_p^2, \dots, \sigma_p^2)$  and

$$Q_\psi = s_0^\psi \begin{bmatrix} 1 & -\frac{\bar{x}_1^\psi}{s_1^\psi} & -\frac{\bar{x}_2^\psi}{s_2^\psi} & \cdots & -\frac{\bar{x}_p^\psi}{s_p^\psi} \\ 0 & \frac{1}{s_1^\psi} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{s_2^\psi} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{s_p^\psi} \end{bmatrix}.$$

It remains to specify  $\sigma_p^2$ . The arguments of Raftery (1996) and subsequent experience using Bayes factors for GLM's (e.g. Viallefont et al. (2001)) suggests using the value  $\sigma_p^2 = 1$ .

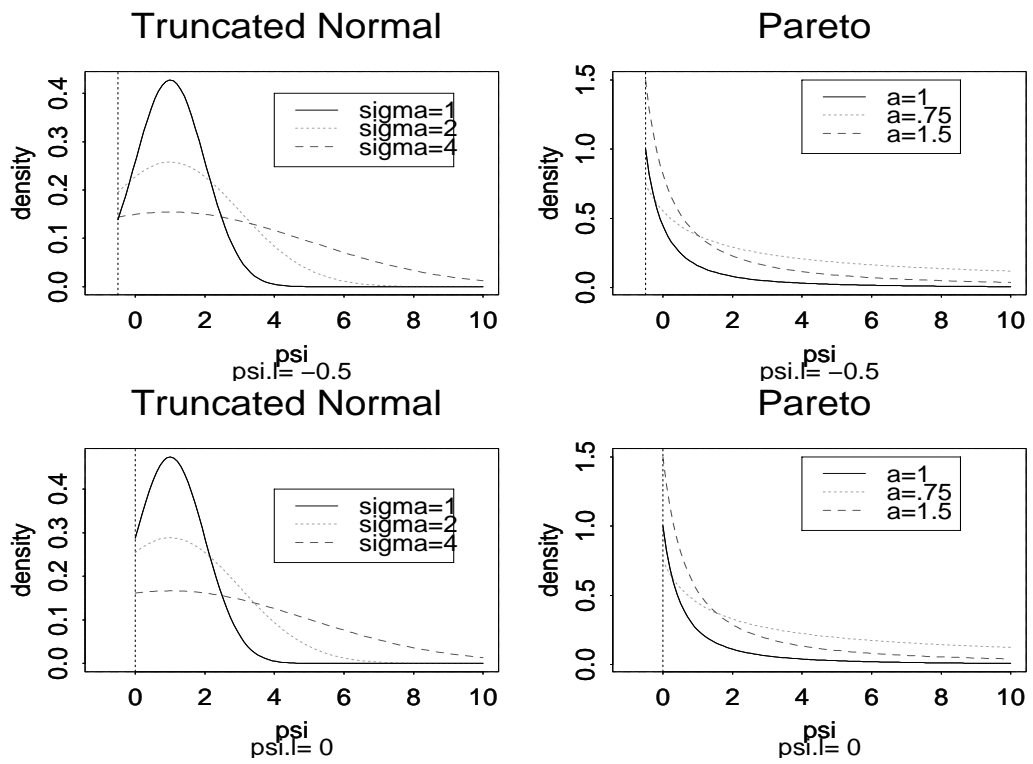
We now consider the choice of the prior distribution for the link parameter  $\psi$ . For the link families specified in (2.3) and (2.4), we require that  $\psi_1 \geq \psi_l$  and  $\psi_2 \geq \psi_l$ , where  $\psi_l$  is chosen in such a way that the restriction on the range of the allowable mean values is reasonable. As noted before for binomial links,  $\psi_l = -.46$  restricts the success probabilities to be between .1 and .9 for  $\psi_1 < 0$  and  $\psi_2 < 0$ . If  $\psi_l = 0$ , there are no restrictions on the success probabilities. Therefore it makes sense to consider prior distributions that are truncated to  $[\psi_l, \infty)$  as prior distributions for  $\psi_1$  and  $\psi_2$ , respectively. As a first choice we consider a truncated normal distribution with mean 1, corresponding to the canonical link, and standard deviation  $\sigma_\psi$ . The left column of Figure 1 shows the corresponding prior densities for  $\psi_l = -.5$  and  $\psi_l = 0$ .

As a second choice we consider Pareto densities given by

$$f(\psi; a, \psi_l) = \frac{a}{(1 + \psi - \psi_l)^{a+1}}, \quad \forall \psi \geq \psi_l, a > 0. \quad (3.9)$$

This prior choice can be motivated by the same arguments as those of Ntzoufras et al. (2001) on page 11. In particular, the link function  $g(\mu; \psi) := F^{-1}(\mu; \psi)$  for (2.3) and (2.4) changes most rapidly for the smallest allowable  $\psi$  value, i.e.  $\psi = \psi_l$ . Therefore Ntzoufras et al. (2001) chose priors which have highest densities at  $\psi_l$  and tails that are monotonically decreasing. The Pareto prior family given in (3.9) satisfies these conditions and is illustrated in the right column of Figure 1.

It seems natural to want the prior distribution of link functions to be centered at, or symmetric about the canonical link function, in some sense, since the canonical link plays a role similar to that of a null hypothesis. As a measure of the symmetry of these prior choices about the canonical link



**Figure 1.**  $N(1, \sigma_\psi)$  priors truncated to  $[\psi_l, \infty)$  and Pareto priors defined in (3.9) for the link parameter  $\psi$ .

**Table 2***P*( $\psi \leq 1$ ) for Pareto and truncated normal prior family for  $\psi$ .

<i>P</i> ( $\psi \leq 1$ ) for Pareto prior			<i>P</i> ( $\psi \leq 1$ ) for $N(1, \sigma_\psi^2)$ truncated to $[\psi_l, \infty)$		
<i>a</i>	$\psi_l = -.5$	$\psi_l = 0$	$\sigma_\psi$	$\psi_l = -.5$	$\psi_l = 0$
.75	.50	.41	1.00	.46	.41
1.00	.60	.50	2.00	.35	.28
1.50	.75	.65	4.00	.23	.16

parameter we use  $P(\psi \leq 1)$ , which is tabulated in Table 2 for the different prior choices considered. For the Pareto priors this shows that  $a = .75$  can be considered symmetric about the canonical link  $\psi = 1$  if  $\psi_l = -.5$ , while  $a = 1$  is symmetric if  $\psi_l = 0$ . For the truncated normal priors, link values larger than 1 are favored when these parameter values are used.

So far we have considered only single tail modifications. For the both tails case (2.2) with  $\boldsymbol{\psi} = (\psi_1, \psi_2)$ , we assume independence of the components and use the same priors for  $\psi_1$  and  $\psi_2$  as for the single tail modifications.

To approximate the Bayes factors  $B_\psi$  of (3.1) we use the Laplace approximation for Bayes factors for GLM's given in Raftery (1996), namely

$$2 \log B_\psi \approx \chi_\psi^2 + (E_\psi - E_0), \quad (3.10)$$

where  $\chi_\psi^2 = dev(M_c) - dev(M_\psi)$ . Here  $dev(M)$  denotes the deviance of model  $M$ . Let  $F_\psi$  denote the observed or expected Fisher information matrix at the MLE  $\hat{\boldsymbol{\beta}}_\psi$  in the model  $M_\psi$ . Then  $E_\psi$  in equation (3.10) is given by

$$E_\psi = \log |G_\psi| - (\hat{\boldsymbol{\beta}}_\psi - \mathbf{v}_\psi)' C_\psi (\hat{\boldsymbol{\beta}}_\psi - \mathbf{v}_\psi) - \log |F_\psi + G_\psi|, \quad (3.11)$$

where  $G_\psi = (Q_\psi U Q_\psi')^{-1}$  is the inverse of the prior variance in (3.6) and  $C_\psi$

is defined as

$$C_\psi = G_\psi \{I - H_\psi(2I - F_\psi H_\psi)G_\psi\}, \text{ where } H_\psi = (F_\psi + G_\psi)^{-1}.$$

Finally,  $E_0$  is equal to  $E_\psi$ , where  $\psi$  is taken to be the value corresponding to the canonical link. Equation (3.11) corresponds to equation (9) in Raftery (1996).

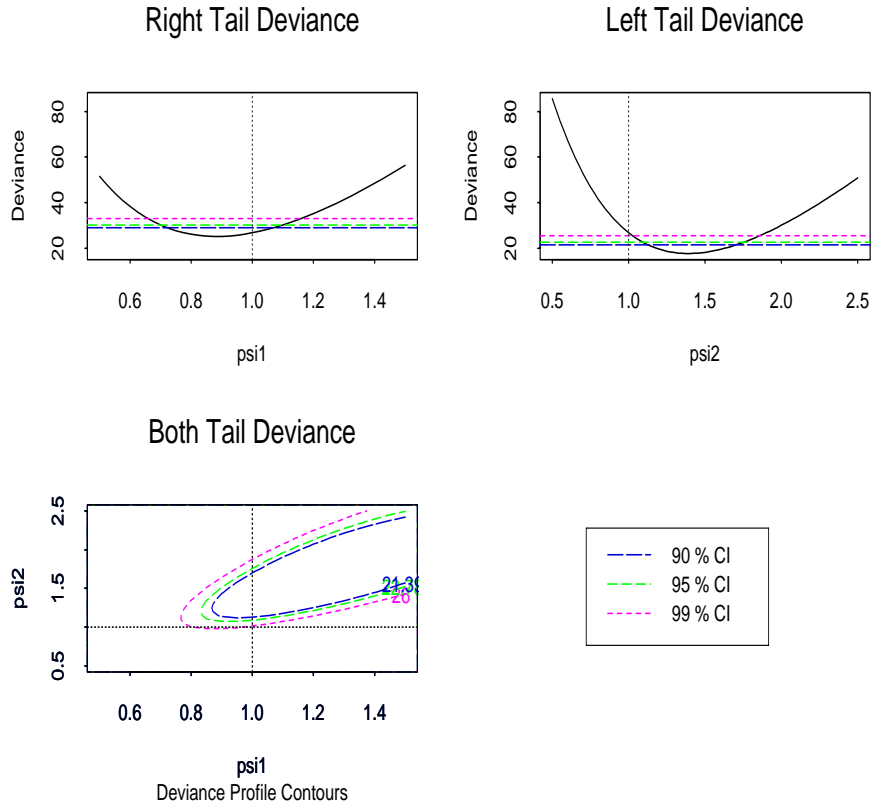
To calculate approximations to the overall Bayes factors specified in (3.3)-(3.5) we use the above approximation and numerically integrate out  $\psi$  using the prior specifications for  $\psi$ .

## 4. Applications

### 4.1 *Age of Menarche in Polish girls*

Milicer and Szcotka (1966) present a sample of 3918 Warsaw girls and record whether or not they had reached menarche together with their age. This is a well known data set and has often been used to demonstrate the need for a link function other than the logistic one. The residual deviance for the logistic regression model with a linear age covariate is 26.70 with 23 degrees of freedom suggesting the possibility of some improving of the fit. Figure 2 gives the deviance profiles and contours when the link families (2.2)-(2.4) are used for binomial regression. The corresponding approximate Bayes factors  $B_\psi$  as a function of  $\psi$  using  $\sigma_p = 1$  are given in Figure 3. The minimal deviances and the maximal approximate Bayes factors are presented in Table 3.

From Table 3 we see that a left tail modification improves the fit using either deviances or Bayes factors. While the deviance indicates that a single tail modification is sufficient, the maximal approximate Bayes factor for the both tail modification is quite high. When looking at the maximal

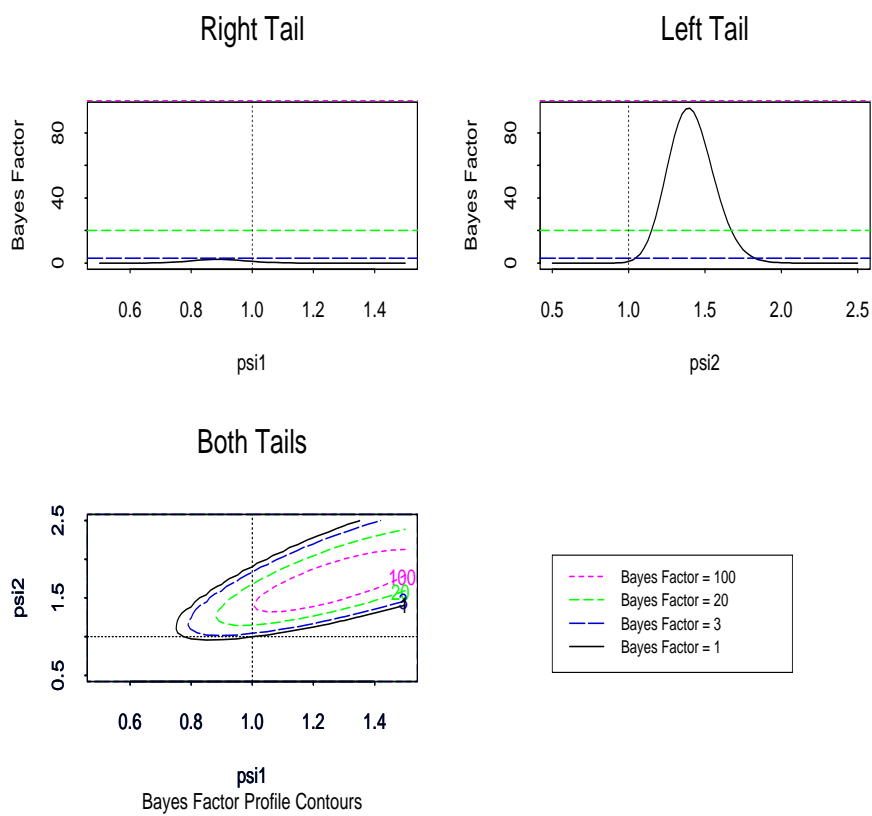


**Figure 2.** Minimal Deviance Profiles and Minimal Deviance Contours for the Menarche Data

**Table 3**

*Minimal Deviances and Approximate Maximal Bayes Factors for the Generalized Link Functions against the Canonical Link, for the Menarche Data (with prior variance parameter  $\sigma_p = 1$ )*

Model	Minimal Deviance	$(\psi_1, \psi_2)$	df	Maximal Bayes Factor	$(\psi_1, \psi_2)$
<b>Right</b>	25.10	(.88, -)	22	2.23	(.75, -)
<b>Left</b>	17.62	(-, 1.40)	22	95.05	(-, 1.40)
<b>Both</b>	15.38	(1.25, 1.67)	21	287.14	(1.27, 1.72)



**Figure 3.** Approximate Bayes Factor Profiles and Contours for the Menarche Data, with prior variance parameter  $\sigma_p = 1$ .



**Table 4**

*Approximate Overall Bayes Factors for the Generalized Link Functions  
Against the Canonical Link, for the Menarche Data with  $\psi_l = 0$*

---



---

$a$	Pareto prior			$\sigma_\psi$	$N(1, \sigma_\psi^2)$ truncated to $[\psi_l, \infty)$		
	Right Tail	Left Tail	Both Tails		Right Tail	Left Tail	Both Tails
.75	.12	5.73	.14	1.00	.23	15.22	9.64
1.00	.13	5.83	.15	2.00	.14	9.93	4.68
1.50	.15	5.97	.15	4.00	.08	5.84	1.69

approximate Bayes factor we ignore the error made by estimating the link parameter and therefore it is more appropriate to consider the overall Bayes factor, which accounts for link uncertainty. Since the observed success probabilities vary between 0 and 1 and since a large range of ages was investigated, it seems reasonable not to impose any restriction on the allowable success probabilities. Therefore we assume  $\psi_l = 0$ . Table 4 gives the overall Bayes factors for the different prior choices for the link parameter.

From Table 4 we see that a left tail modification improves the fit regardless of which prior specification is used for the link parameter and if we account for link uncertainty. The approximate overall Bayes factors also show that a right tail modification gives no improvement for all priors used, while there is slight evidence for a both tail modification when truncated normal priors are used. Note that these priors give more probability to link values larger than the canonical link value. From Figure 3 we see that  $B_\psi$  is largest for link values greater than 1. In contrast, the Pareto priors favor small link values. This explains that the approximate overall Bayes factors are lower for the Pareto priors compared to the truncated normal priors.

Before we can conclude that a link function with a left tail modification is better than the logistic link for these data, we need to consider and exclude possible alternative explanations of what we have observed. It is possible that a transformation of the age variable may be preferable to a link modification, and that the apparently poorer performance of the logistic link is just an artifact due to nonlinearity of the effect of age. We now investigate this possibility.

We consider polynomial models for age of the form

$$\eta_i = \beta_0 + \beta_1 age_i + \cdots + \beta_p age_i^p.$$

Calculations of the deviances show that the quadratic model is little better than the linear one, that there is a substantial reduction in deviance when one goes from quadratic to cubic, and little further gain for additional polynomial terms. We therefore restrict ourselves to considering a cubic model for age.

Our question is thus whether a left tail modification with linear age is better than a canonical link with cubic age. The most common approach to this question is to compute deviances and degrees of freedom for the two competing models. The left tail modification with linear age has deviance 17.62 on 22 d.f., while the canonical link with cubic age has deviance 15.04 on 21 d.f. These models are not nested, so a standard likelihood ratio test cannot be carried out. Nevertheless, the deviance difference is 2.58 with a difference in degrees of freedom of 1, and if this was compared with the standard chi-squared distribution (which cannot validly be done), it would not be significant, and one would typically choose the more parsimonious left tail modification with linear age model.

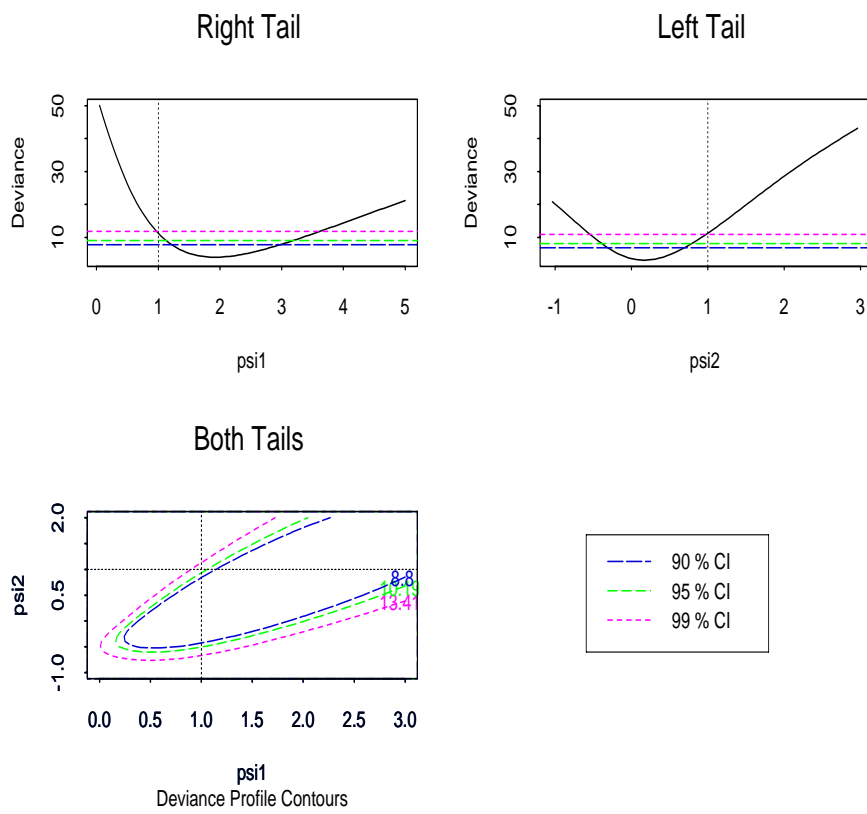
**Table 5**  
*Beetle Mortality Data*

$Y_i$ Number killed	$n_i$ Number of Insects	Dose $\log_{10} CS_2 mgl^{-1}$
6	59	1.6907
13	60	1.7242
18	62	1.7552
28	56	1.7842
52	63	1.8113
53	59	1.8369
61	62	1.8610
60	60	1.8839

Bayes factors do allow us to make a formal comparison between these two nonnested models. The Bayes factor for the left tail modification with linear age model against the canonical link model with cubic age is 233 for the Pareto prior with  $a = 1$ , and 396 for the truncated normal prior with  $\sigma = 2$ , so that the left tail modification with linear age model is favored. Thus with Bayes factors we reach the same conclusion as with the informal comparison of deviances, with the difference that Bayes factors provide a formal justification for the conclusion.

#### 4.2 *Beetle Mortality*

Bliss (1935) recorded the number of insects dead after five hours' exposure to gaseous carbon disulphide at various concentrations and the data are presented in Table 5. This is also a well known data set for investigating a different link function other than the logistic one. Here, the residual deviance for a logistic model with a centered log dose covariate is 11.23 with 6 degrees of freedom, suggesting some lack of fit.



**Figure 4.** Deviance Profiles and Deviance Contours for the Beetle Mortality Data

**Table 6**

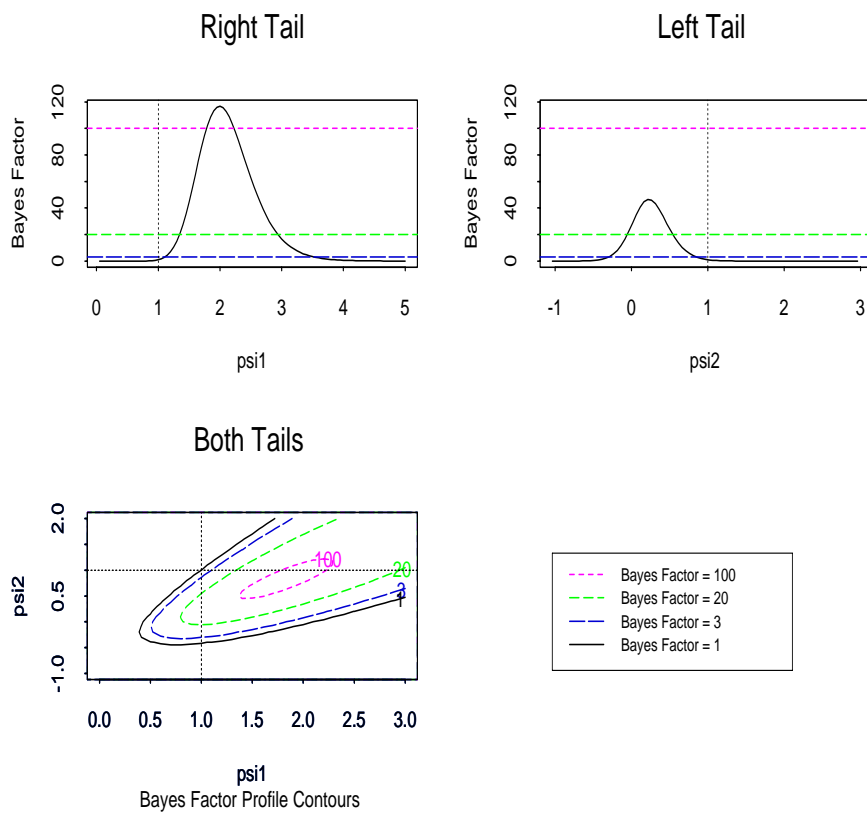
*Minimal Deviances and Approximate Maximal Individual Bayes Factors for the Beetle Mortality Data (prior variance parameter  $\sigma_p = 1$ )*

<b>Model</b>	<b>Minimal Deviance</b>	$(\psi_1, \psi_2)$	<b>df</b>	<b>Maximal Bayes Factor</b>	$(\psi_1, \psi_2)$
<b>Right</b>	3.96	(1.92, -)	5	116.66	(1.99, -)
<b>Left</b>	3.04	(-, .16)	5	46.41	(-, .21)
<b>Both</b>	2.81	(1.2, .3)	4	123.89	(1.8, .8)

Figure 4 gives the deviance profiles and contours, when the link families (2.2)-(2.4) are used for binomial regression. They suggest that a link tail modification in this data set is useful and improves the fit. We will now use Bayes factors to decide which specific tail modification is needed. We use the prior specification (3.6) with  $\sigma_p = 1$  for the regression coefficients. Figure 5 shows the Bayes factors  $B_\psi$  as a function of  $\psi$ , and in Table 6 we give the minimal deviances and maximal individual Bayes factors  $B_\psi$  for each tail modification.

From this we conclude that the Bayes factors clearly favor a right tail or both tail modification over a left tail modification. While the likelihood ratio test can be used to show that the reduction in deviance achieved by using a both tail modification over a right/or left tail modification is insignificant, we cannot compare right and left tail modifications, since they are not nested models. Graphically, we see that in Figure 4, the lines determining the point (1,1) (corresponding to logistic link) intersect the confidence regions, suggesting that single tail modifications are sufficient.

We now take into account the link uncertainty by considering overall



**Figure 5.** Approximate Bayes Factor Profiles and Contours for the Beetle Mortality Data with  $\sigma_p = 1$

**Table 7***Approximate Overall Bayes Factors for the Beetle Data with  $\psi_l = -.5$* 

$a$	Pareto prior			$N(1, \sigma_\psi^2)$ truncated to $[\psi_l, \infty)$			
	Right Tail	Left Tail	Both Tails	$\sigma_\psi$	Right Tail	Left Tail	Both Tails
.75	10.10	8.08	.59	1.00	27.76	8.75	15.37
1.00	9.80	9.48	.60	2.00	26.65	6.55	10.83
1.50	7.77	11.00	.90	4.00	18.14	4.15	5.28

Bayes factors, which are given in Table 7 for  $\psi_l = -.5$ . We use  $\psi_l = -.5$  since the observed success probabilities vary between .1 and .9. Therefore we would like to allow for links which take this restriction into account. For the Pareto priors a single tail modification is sufficient, but the difference between a left tail or right tail is minimal. The truncated normal priors favor a right tail modification over a left tail or both tail modification. The difference in the results for the two prior specifications can be explained as follows. Pareto priors favor small link values while truncated normal priors favor large link values. In this data set, this corresponds to left tail modifications for the Pareto priors and right tail modifications for the truncated normal priors.

This data set has also been considered by Collett (1991) p. 108-112, who allowed for the inclusion of a quadratic term on the original  $\mathbf{CS}_2$  scale in a logistic model. This yields a residual deviance of 3.08 with 5 degrees of freedom. We can now use Bayes factors to decide if the right tail link fit is preferable over the inclusion of a quadratic term on the original  $\mathbf{CS}_2$  scale. Note that these models are again nonnested. The corresponding Bayes factor

is given by

$$B_{\psi_1=1.99} \times \frac{Pr(\mathbf{Y}|M_{\psi_1=1,x=\log(CS_2)})}{Pr(\mathbf{Y}|M_{\psi_1=1,x=(CS_2,CS_2^2)})} = 116.66 \times .0011 = .1280 = \frac{1}{7.80},$$

and the overall Bayes factor using a right tail modification

$$B_r \times \frac{Pr(\mathbf{Y}|M_{\psi_1=1,x=\log(CS_2)})}{Pr(\mathbf{Y}|M_{\psi_1=1,x=(CS_2,CS_2^2)})}$$

varies between  $\frac{1}{32}$  and  $\frac{1}{117}$  for the different link prior specifications. This shows that a logistic model using a quadratic term on the original scale is favored over a right tail link family. Collett (1991) p. 140 noted that a complementary log-log model for the link parameter fits the data as well as the logistic model using a quadratic term. He argued that the complementary log-log model would be preferable since it has fewer parameters, but this ignores the uncertainty in the choice of link function.

## 5. Discussion

We have presented a Bayesian approach to model selection in GLM's with parametric link using Bayes factors to account for structural model uncertainty (see Draper (1995)) such as the choice of link in a GLM. This involves a continuous model expansion over ordinary GLM's when a particular link family was considered as well as a discrete model expansion when different link families were compared. In addition we were able to jointly assess the choice of link together with the choice of the set of independent parameters to include in the model. This involves the comparison of nonnested models, which cannot be carried out using classical model selection strategies based on significance tests.

We used reference proper priors for the regression parameters of a GLM with a fixed link function as suggested by Raftery (1996). These priors vary



with the link parameter, reflecting the fact that the regression parameters are on different scales for different link functions. This reference proper prior avoids the problem of Bartlett's (Bartlett (1957)) or Lindley's (Lindley (1957)) paradox and thus in this case Bayes factors have the advantage over posterior Bayes factors (Aitkin (1991)),  $p$ -values or the AIC criterion that they correctly identify the correct model in large samples, while the other criteria do not (Schwartz (1978)). Different prior distributions of the link parameter were investigated. Finally, the Bayes factors were approximated using the Laplace approximations given in Raftery (1996). With regard to prior sensitivity we observed the following. For the large menarche data set the qualitative conclusions are unchanged by the link prior specification, while for the smaller beetle data set we observed a moderate dependency on the prior specification. This kind of behavior is common in Bayesian analysis.

A complete Bayesian analysis of a GLM with a parametric link is computer intensive, since the calculation of posterior distributions involves Markov Chain Monte Carlo methods (Czado (1994)). The methods presented in this paper can be used for a final analysis, or could be used to screen for plausible models, which could then be used as starting points for a complete Bayesian analysis. Note that our methods for calculating these Bayes factors only require software that is able to fit a GLM with an arbitrary link. In particular, joint maximization over regression parameters and link parameters to determine the maximum likelihood estimator is not needed. Here, calculations were conducted in S-Plus using the `glm()` function together with integration functions in one or two dimensions.

It should be noted that Bayes factors address the issue of model choice,

and not parameter estimation. For inference about model-independent quantities such as the log odds ratio of a treatment effect or the mean response at a particular value of the independent variables, methods for taking account of model uncertainty such as Bayesian model averaging (see for example Hoeting et al. (1999)) are needed. This also allows a Bayesian alternative to the quantifications of change to quantities of interest when changing from a GLM with canonical link to one with noncanonical link. This was the goal of a paper by Czado and Munk (2000).

#### ACKNOWLEDGEMENTS

The first author was supported by Sonderforschungsbereich 386 *Statistische Analyse Diskreter Strukturen*, and the second author by NIH Grant 1R01CA094212-01 and ONR Grant N00014-01-10745.

#### REFERENCES

- Aitkin, M. (1991). Posterior Bayes factors (disc: P128-142). *Journal of the Royal Statistical Society, Series B, Methodological* **53**, 111–128.
- Aranda-Ordaz, F. J. (1981). On two families of transformations to additivity for binary response data (corr: V70 p303). *Biometrika* **68**, 357–363.
- Bartlett, M. S. (1957). A comment on D. V. Lindley’s statistical paradox. *Biometrika* **44**, 533.
- Bliss, C. (1935). The calculation of the dose-mortality curve. *Annals of Applied Biology* **22**, 134–167.
- Collett, D. (1991). *Modelling Binary Data*. Chapman & Hall.

- Copenhaver, T. W. and Mielke, P. W. (1977). Quantit analysis: A quantal assay refinement. *Biometrics* **33**, 175–186.
- Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (c/r: P18-39). *Journal of the Royal Statistical Society, Series B, Methodological* **49**, 1–18.
- Czado, C. (1992). On link selection in generalized linear models. In *Advances in GLIM and Statistical Modelling. Proceedings of the GLIM92 Conference*, pages 60–65. Springer-Verlag (Berlin; New York).
- Czado, C. (1994). Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference* **41**, 121–140.
- Czado, C. (1997). On selecting parametric link transformation families in generalized linear models. *Journal of Statistical Planning and Inference* **61**, 125–139.
- Czado, C. and Munk, A. (2000). Noncanonical links in generalized linear models - when is the effort justified? *Journal of Statistical Planning and Inference* **87**, 317–345.
- Czado, C. and Santner, T. J. (1992). The effect of link misspecification on binary regression inference. *Journal of Statistical Planning and Inference* **33**, 213–231.
- Draper, D. (1995). Assessment and propagation of model uncertainty (disc: P71-97). *Journal of the Royal Statistical Society, Series B, Methodological* **57**, 45–70.
- Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B, Methodological* **56**, 501–514.

- Guerrero, V. M. and Johnson, R. A. (1982). Use of the Box-Cox transformation with binary response models. *Biometrika* **69**, 309–314.
- Han, C. and Carlin, B. (2001). MCMC methods for computing Bayes factors: a comparative review. *Journal of the American Statistical Association* **96**, 1122–1132.
- Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial (disc: P401-417). *Statistical Science* **14**, 382–401. Corrected version available at [www.stat.washington.edu/www/research/online/hoeting1999.pdf](http://www.stat.washington.edu/www/research/online/hoeting1999.pdf).
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.
- Lindley, D. V. (1957). On the presentation of evidence. *Biometrika* **44**, 187–192.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London, 2nd edition.
- Milicer, H. and Szcotka, F. (1966). Age at menarche in Warsaw girls in 1965. *Human Biology* **40**, 199–203.
- Montfort, M. A. J. and Otten, A. (1976). Quantal response analysis: Enlargement of the logistic model with a kurtosis parameter. *Biometrical Journal* **18**, 371–380.
- Morgan, B. J. T. (1983). Observations on quantit analysis. *Biometrics* **39**, 879–886.
- Ntzoufras, I., Dellaportas, P. and Forster, J. J. (2001). Bayesian variable and link determination for generalised linear models. *preprint* .
- Pregibon, D. (1980). Goodness of link tests for generalized linear models.

- Applied Statistics* **29**, 15–24.
- Prentice, R. L. (1976). Generalization of the probit and logit methods for dose response curves. *Biometrics* **32**, 761–768.
- Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika* **83**, 251–266.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–64.
- Stukel, T. (1988). Generalized logistic models. *Journal of the American Statistical Association* **83**, 426–431.
- Taylor, J. M. G. (1988). The cost of generalizing logistic regression. *Journal of the American Statistical Association* **83**, 1078–1083.
- Taylor, J. M. G., Siqueira, A. L. and Weiss, R. E. (1996). The cost of adding parameters to a model. *Journal of the Royal Statistical Society, Series B, Methodological* **58**, 593–607.
- Viallefont, V., Raftery, A. E. and Richardson, S. (2001). Variable selection and Bayesian model averaging in epidemiological case-control studies. *Statistics in Medicine* **20**, 3215–3230.
- Whittmore, A. (1983). Transformations to linearity in binary regression. *SIAM Journal on Applied Mathematics* **43**, 703–710.