

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

Claudia Czado*

Zentrum Mathematik, Technische Universität München,
Arcistr. 21, D-80290 Munich, Germany

and

Adrian E. Raftery†

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

Technical Report no. 401
Department of Statistics
University of Washington

October 16, 2001

SUMMARY. One important component of model selection using generalized linear models (GLM) is the choice of a link function. Approximate Bayes factors are used to assess the improvement in fit over a GLM with canonical link when a parametric link family is used. For this approximate Bayes factors are calculated using the approximations given in Raftery (1996), together with a reference set of prior distributions. This methodology can also be used to differentiate between different parametric link families, as well as allowing one to jointly select the link family and the independent variables. This involves comparing nonnested models. This is illustrated using parametric link families studied in Czado (1997) for two data sets involving binomial responses.

KEYWORDS: Bayes factor; Generalized linear model; Link function; Model selection; Reference prior.

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

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

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 16 OCT 2001		2. REPORT TYPE		3. DATES COVERED 00-10-2001 to 00-10-2001	
4. TITLE AND SUBTITLE Choosing the Link Function and Accounting for Link Uncertainty in Generalized Linear Models using Bayes Factors				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Washington, Department of Statistics, Box 354322, Seattle, WA, 98195-4322				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 24	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Contents

1	Introduction	3
2	Generalized Linear Models with Parametric Links	4
3	Approximate Bayes Factors for GLM's with Parametric Link	5
4	Applications	9
4.1	Beetle Mortality	9
4.2	Rotifer Suspension	10
5	Discussion	11

List of Tables

1	Beetle Mortality Data	22
2	Approximate Bayes Factors and Deviances for the Beetle Mortality Data ($\sigma_p = 1, \sigma_\psi = 2$)	23
3	Approximate Log Bayes Factors and Deviances for the Rotifer Suspension Data ($\sigma_p = 1, \sigma_\psi = 2$)	24

List of Figures

1	Deviance Profiles and Deviance Contours for the Beetle Mortality Data . . .	16
2	Approximate Bayes Factor Profiles and Contours for the Beetle Mortality Data with $\sigma_p = 1$	17
3	Deviance Profiles and Deviance Contours using no Interaction Term for the Rotifer Suspension Data	18
4	Deviance Profiles and Deviance Contours using an Interaction Term for the Rotifer Suspension Data	19
5	Approximate Bayes Factors Profiles and Contours using no Interaction Term for the Rotifer Suspension Data with $\sigma_p = 1$	20
6	Approximate Bayes Factors Profiles and Contours using an Interaction Term for the Rotifer Suspension Data with $\sigma_p = 1$	21

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 have to be investigated and compared using individual significance tests based on the asymptotic distribution of the deviance. As pointed out in Gelfand and Dey (1994) and Raftery (1996) this strategy cannot be used for comparing nonnested models. In addition, adjustments for multiple tests as well as power considerations are usually ignored. A Bayesian approach can avoid these difficulties and therefore 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 assess jointly the choice of the link family and of the set of independent variables.

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 ϕ 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 μ_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} which 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 ϕ in (2.1) is typically estimated by an appropriate moment estimator involving the Pearson χ^2 Statistic (McCullagh and Nelder (1989)). For a fixed link parameter ψ we remain in the class of GLM's, while this is no longer true if the link parameter ψ 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.

As in the case for Box-Cox transformations (Box and Cox (1964)) one has to decide whether to make inference conditionally on an estimated link parameter or not. In the Box-Cox controversy, Hinkley and Runger (1984) and Box and Cox (1982) argued for following a conditional approach, while Bickel and Doksum (1981) and Carroll and Ruppert (1981) advocated an unconditional approach. We follow the arguments given in Draper (1995) and are interested in assessing model uncertainty.

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 follows:

Error Distribution	Parameter Restriction	Canonical Link	Link Family $\mathfrak{S} = \{F(\cdot, \psi) : \psi \in \Psi\}$
Normal	μ 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	$\mu > 0$	$F(\eta) = \eta^{-.5}$	$F(\eta, \psi) = [\exp(h(\eta, \psi))]^{-.5}$

where $h(\eta, \psi)$ is one of the following functions:

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

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

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

It should be noted that 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 $\boldsymbol{\psi} = (1, \psi_2)$. Similarly (2.3) is a special case of (2.2) with $\boldsymbol{\psi} = (\psi_1, 1)$. As ψ_1 increases the right tail of $G(\cdot, \boldsymbol{\psi})$ becomes lighter, while an increasing ψ_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 $\psi_1 \neq \psi_2$.

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_ψ a GLM with a fixed link parameter ψ 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_ψ by β_ψ to indicate that the regression parameters are on different scales for different ψ 's. We are interested in the Bayes factor for model M_ψ 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),

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

where β_ψ is the corresponding regression parameter in Model M_ψ and $p(\beta_\psi|M_\psi)$ is its prior density in model M_ψ . Note that M_c corresponds to M_ψ with $\psi = 1$.

The Bayes factor is a summary of the evidence for M_ψ 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. In this paper we follow Raftery (1996) by using the rounded scale given in Table 1 of Raftery (1996) for interpreting B_ψ or $2 \log B_\psi$.

This approach allows us to compare different parametric link families as follows. Let M_θ denote a GLM using a link family indexed by the link parameter θ and construct B_θ in a similar fashion as B_ψ . The quantity $\frac{B_\psi}{B_\theta}$ then provides a summary of the evidence for model M_ψ against model M_θ 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 $\boldsymbol{\psi}$, ψ_1 and ψ_2 , respectively. If the link parameter values are not chosen in advance,

but instead are estimated, B_{ψ_1}, B_{ψ_2} and B_{ψ} will tend to overstate the evidence for a modification. The average 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 Bayes factors, we have to select appropriate 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 to construct overall Bayes factors for a GLM with a specified link family.

For the prior distribution of the regression parameters β_ψ in the model M_ψ we use the reference proper prior distributions suggested by Raftery (1996) for GLM's, since for fixed values of the link parameter ψ 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 β_ψ 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 Σ . 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^ψ (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 ψ , 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 $\mathbf{v}'_\psi = (\bar{z}_\psi, 0, \cdots, p)'$, U denotes a diagonal matrix with

diagonal entries given by $(1, \sigma_p^2, \dots, \sigma_p^2)$ and

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

It remains to specify σ_p^2 . The arguments of Raftery (1996) and subsequent experience using Bayes factors for GLM's (e.g. Viallefont et al. (1998)) suggests using the value $\sigma_p^2 = 1$.

For the prior distribution of the link parameter ψ we use a normal prior centered at the link parameter value corresponding to the canonical link. In particular, for the link family with specification (2.3) or (2.4) we use a normal prior with mean 1 and standard deviation σ_ψ , while for the bivariate specification (2.2) involving $\boldsymbol{\psi} = (\psi_1, \psi_2)$, we assume independence of the components and proceed as in the univariate specifications. Numerical experience with this particular link family suggests that $\sigma_\psi = 2$ is a reasonable choice.

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

$$2 \log B_\psi \approx \chi_\psi^2 + (E_\psi - E_0), \quad (3.9)$$

where $\chi_\psi^2 = dev(M_c) - dev(M_\psi)$. Here $dev(M)$ denotes the deviance of model M . Let F_ψ denote the observed or expected Fisher information matrix at the MLE $\hat{\boldsymbol{\beta}}_\psi$ in the model M_ψ . Then E_ψ in equation (3.9) is given by

$$E_\psi = \log |G_\psi| - (\hat{\boldsymbol{\beta}}_\psi - \mathbf{v}_\psi)' C_\psi (\hat{\boldsymbol{\beta}}_\psi - \mathbf{v}_\psi) - \log |F_\psi + G_\psi|, \quad (3.10)$$

where $G_\psi = (Q_\psi U Q_\psi')^{-1}$ is the inverse of the prior variance in (3.6) and C_ψ 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_ψ , where ψ is taken to be the value corresponding to the canonical link. Equation (3.10) 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 ψ using the prior specifications for ψ .

4. Applications

4.1 *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 2. This is a well known data set and has been often used to show the usefulness of a different link function other than the logistic one. In particular, the residual deviance for a logistic model with a centered log dose covariate is 11.23 with 6 degrees of freedom, suggesting a lack of fit.

[Table 1 about here.]

[Figure 1 about here.]

Figure 1 gives the deviance profiles and contours, when the link families (2.2)-(2.4) are used for binomial regression. They show clearly that a 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 $\sigma_p = 1$ and normal independent priors for $\boldsymbol{\psi}$ with prior standard deviation $\sigma_\psi = 2$. Figure 2 shows the Bayes factors B_ψ as a function of ψ and in Table 3 the overall Bayes factors for each tail modification family are given together with minimal deviances and maximal individual Bayes factors B_ψ .

[Figure 2 about here.]

[Table 2 about here.]

From this we conclude that the Bayes factors clearly favors a right tail modification over a left or both 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 1, the lines determining the point (1,1) (corresponding to logistic link) intersect the confidence regions suggesting that single tail modifications are sufficient. We can also see from Table 3 that the maximal Bayes factors, corresponding to estimated values of ψ_1 and ψ_2 , overestimate the evidence for a modification quite substantially.

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 CS_2 scale. Note these models are again nonnested. The corresponding Bayes factor is given by

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

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.

4.2 *Rotifer Suspension*

The following example is taken from Collett (1991). It involves the number of rotifers falling out of suspension for two species, called *polyartha major* and *keratella cochlearis* for different fluid densities; the data are given in Table 6.10 in Collett (1991), p. 217. For the binary regression models considered below species were coded by 1 for *polyartha major* and 0 for *keratella cochlearis* and a centered covariate for density \times 100 was used.

In this data set we have in addition to the link choice the problem of deciding whether or not to include an interaction term between species and density. A logistic regression analysis gives a residual deviance of 434.25 on 37 degrees of freedom for a model including no interaction term, while a model including an interaction term yields a residual deviance of 434.01 on 36 degrees of freedom. This indicates a severe lack of fit and shows that an interaction term is not needed if only a logistic link is allowed. Therefore, this raises the question whether an interaction term would improve the fit when links other than the logistic are considered. So we consider six model classes corresponding to the three possible tail modifications and the two choices for set of covariates. Figures 3 and 4 present the corresponding deviance profiles and contours. This suggests that the inclusion of an interaction term decreases the residual deviance substantially. Further, a both tail modification substantially improves the fit, which can be seen since the lines determining the logistic link ($\psi = (1, 1)$) do not intersect with the confidence regions. This was also noted by Czado (1994b) who conducted a fully Bayesian analysis of this data set using Markov Chain Monte Carlo methods for joint inference on regression parameters and link parameters.

[Figure 3 about here.]

[Figure 4 about here.]

Figures 5 and 6 present the approximate Bayes factors B_ψ as a function of ψ , and in Table 4 the overall Bayes factors for each tail modification family and covariate set are given together with minimal deviances and maximal individual Bayes factors B_ψ :

[Figure 5 about here.]

[Figure 6 about here.]

[Table 3 about here.]

The overall log Bayes factors for the both tail modification is the largest for the model including an interaction term. Comparing whether the inclusion of an interaction term is warranted for a both tail modification we can see that the log Bayes factor for an interaction is $76.84 - 67.24 = 9.6$, which corresponds to strong evidence for an interaction. Note that a conventional GLM analysis such as that of Collett (1991) would miss this important interaction.

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)). Finally, the Bayes factors were approximated using the Laplace approximations given in Raftery (1996).

A complete Bayesian analysis of a GLM's with a parametric link is computer intensive, since the calculation of posterior distributions involve Markov Chain Monte Carlo methods (Czado (1994a)). 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 our methods for calculating these Bayes factors only requires software which 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 give summary statistics for the fit of a particular model or model class. 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 techniques such as Bayesian model averaging (see for example Hoeting et al. (1999)) or MCMC methods to compute posterior quantities are required. 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 ONR Grant no. N0014-96-1-1092.

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.
- Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association* **76**, 296–311.
- Bliss, C. (1935). The calculation of the dose-mortality curve. *Annals of Applied Biology* **22**, 134–167.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, Methodological* **26**, 211–252.
- Box, G. E. P. and Cox, D. R. (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association* **77**, 209–210.
- Carroll, R. J. and Ruppert, D. (1981). On prediction and the power transformation family. *Biometrika* **68**, 609–615.
- 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. (1994a). Bayesian inference of binary regression models with parametric link. *Journal of Statistical Planning and Inference* **41**, 121–140.
- Czado, C. (1994b). Parametric link modification of both tails in binary regression. *Statistical Papers* **35**, 189–201.
- 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.
- Hinkley, D. V. and Runger, G. (1984). The analysis of transformed data (c/r: P309-320). *Journal of the American Statistical Association* **79**, 302–309.
- 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.
- 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.
- 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 quantal analysis. *Biometrics* **39**, 879–886.
- 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*

- tion **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. (1998). Variable selection and Bayesian model averaging in case-control studies. *Technical Report no. 343, Dept. Of Statistics, University Of Washington* .
- Whittmore, A. (1983). Transformations to linearity in binary regression. *SIAM Journal on Applied Mathematics* **43**, 703–710.

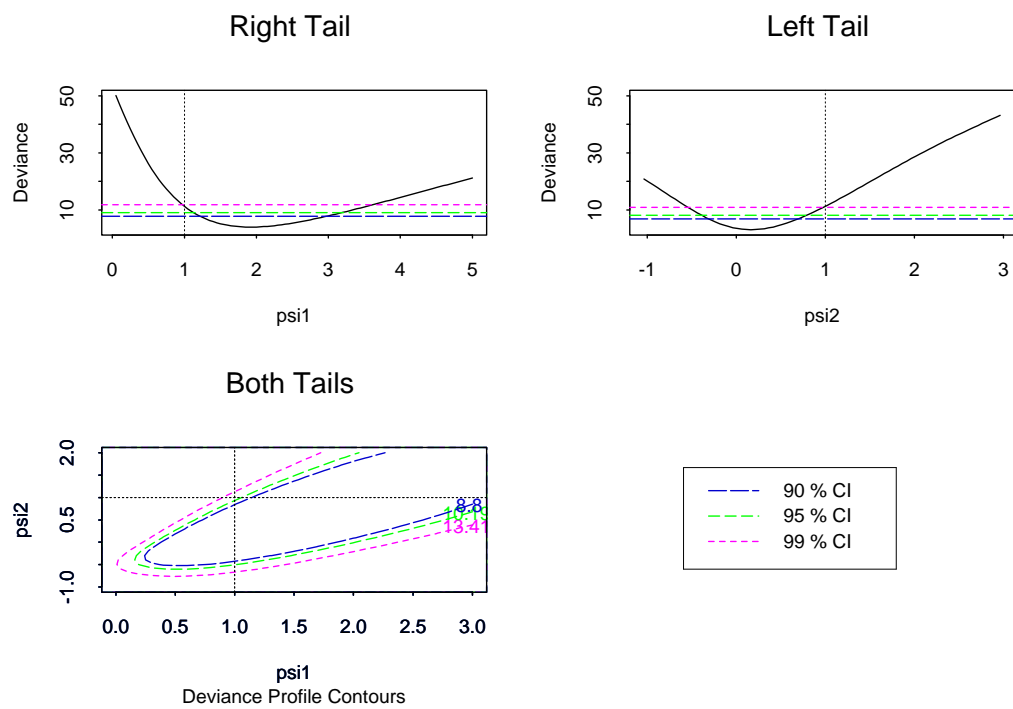


Figure 1. Deviance Profiles and Deviance Contours for the Beetle Mortality Data

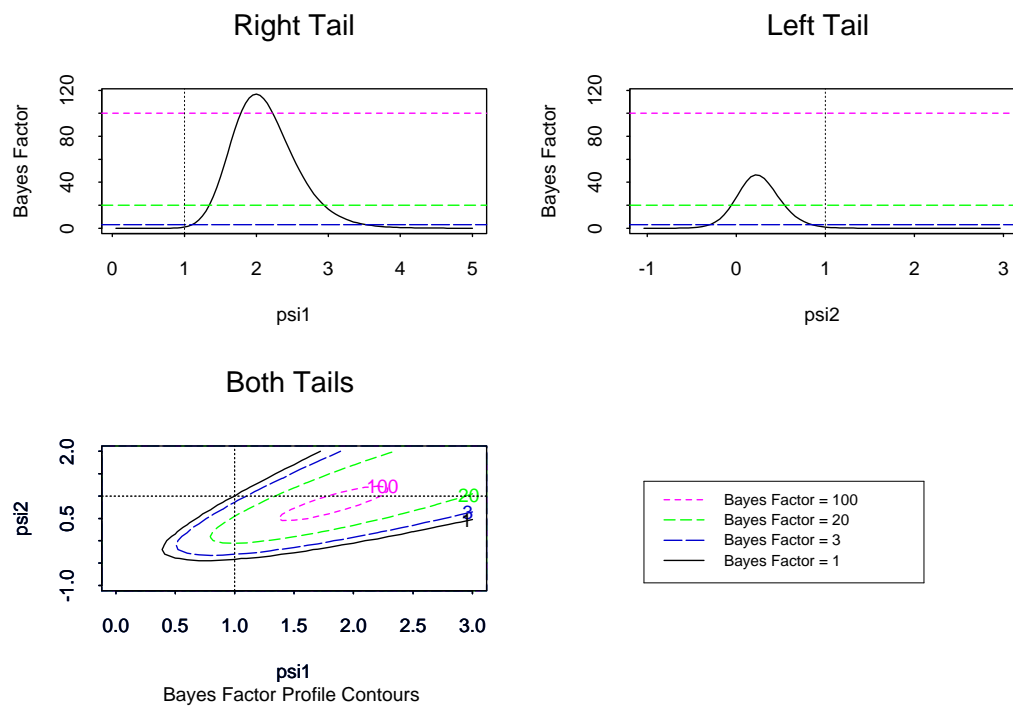


Figure 2. Approximate Bayes Factor Profiles and Contours for the Beetle Mortality Data with $\sigma_p = 1$

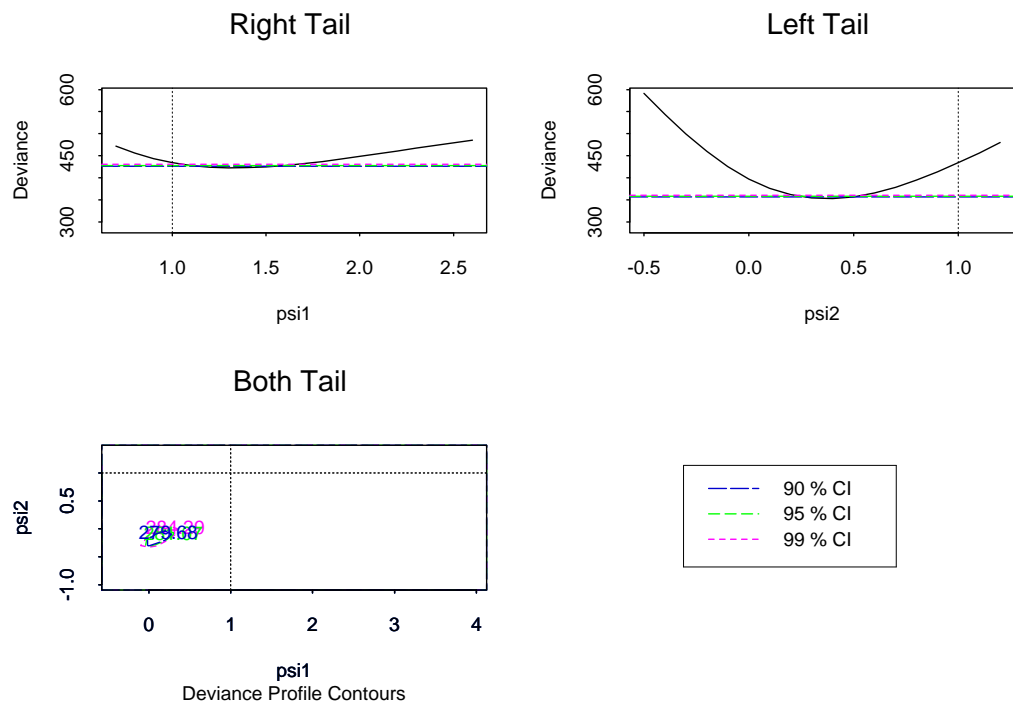


Figure 3. Deviance Profiles and Deviance Contours using no Interaction Term for the Rotifer Suspension Data

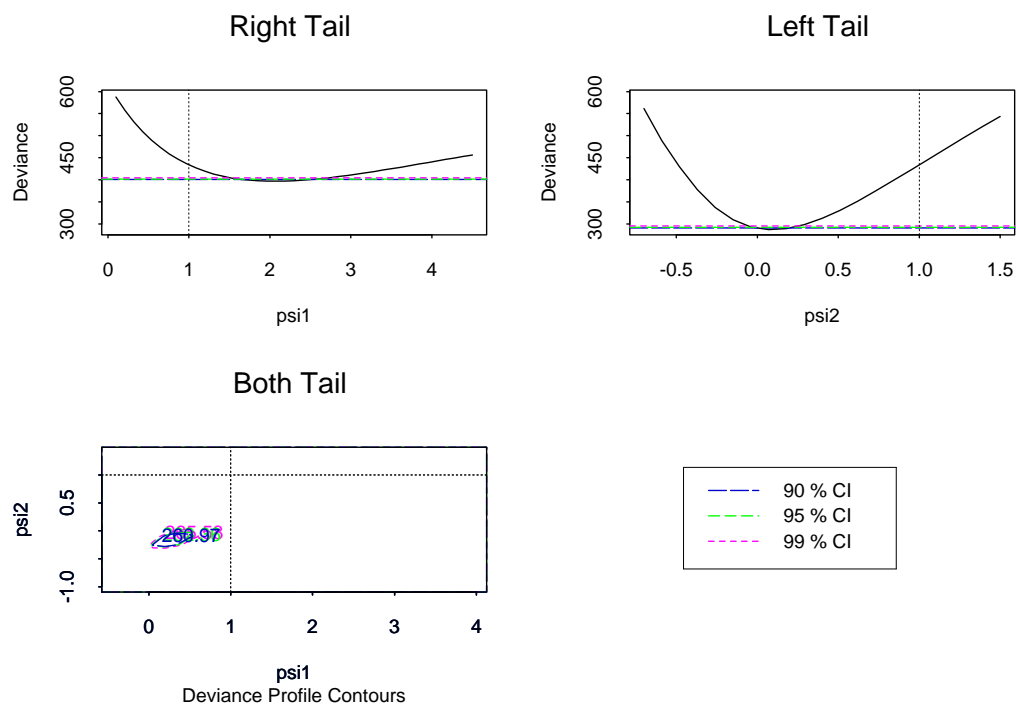


Figure 4. Deviance Profiles and Deviance Contours using an Interaction Term for the Rotifer Suspension Data

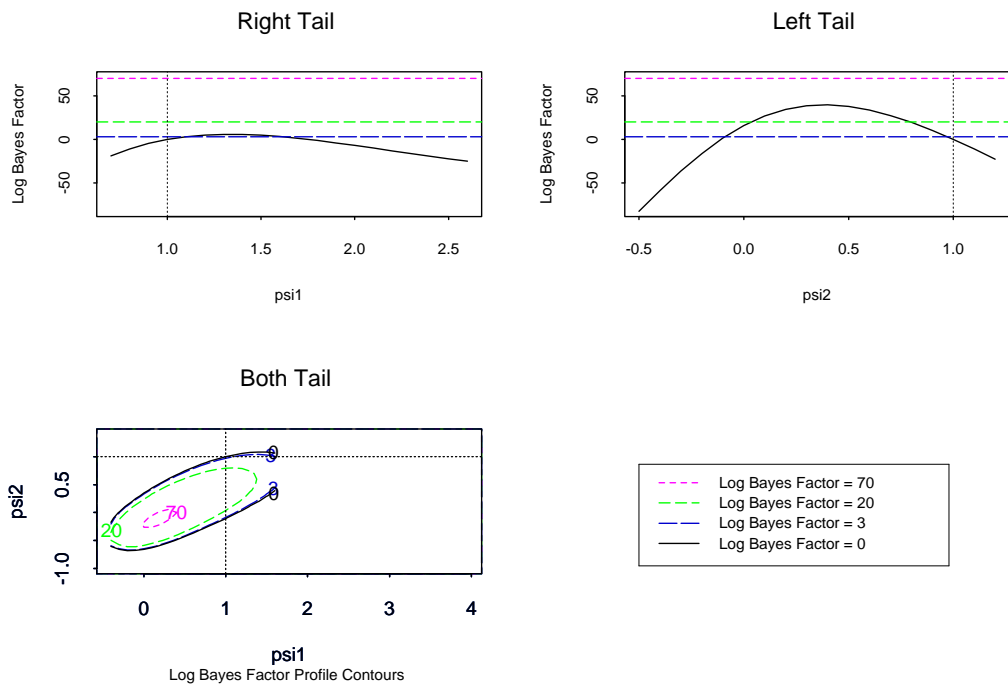


Figure 5. Approximate Bayes Factors Profiles and Contours using no Interaction Term for the Rotifer Suspension Data with $\sigma_p = 1$

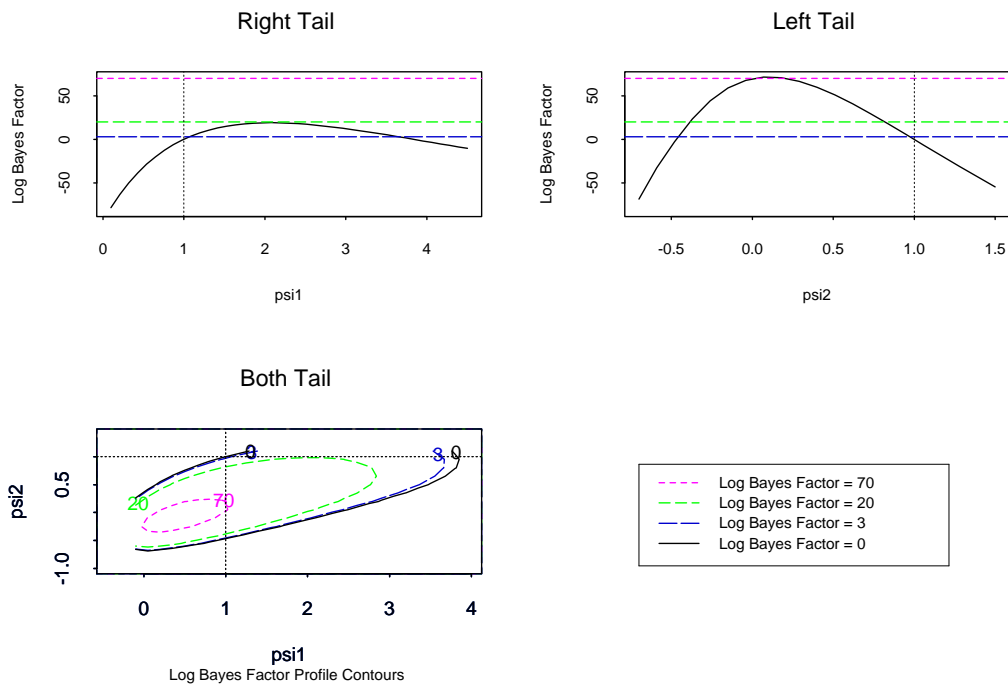


Figure 6. Approximate Bayes Factors Profiles and Contours using an Interaction Term for the Rotifer Suspension Data with $\sigma_p = 1$

Table 1
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

Table 2

Approximate Bayes Factors and Deviances for the Beetle Mortality Data ($\sigma_p = 1, \sigma_\psi = 2$)

Model	Minimal Deviance	(ψ_1, ψ_2)	df	Maximal Bayes Factor	(ψ_1, ψ_2)	Overall Bayes Factor
Right	3.96	(1.92, -)	5	116.66	(1.99, -)	20.61
Left	3.04	(-, .16)	5	46.41	(-, .21)	5.07
Both	2.81	(1.2, .3)	4	123.89	(1.8, .8)	5.38

Table 3
Approximate Log Bayes Factors and Deviances for the Rotifer Suspension Data
 $(\sigma_p = 1, \sigma_\psi = 2)$

Model	Minimal Deviance	(ψ_1, ψ_2)	df	Maximal Log Bayes Factor	(ψ_1, ψ_2)	Overall Log Bayes Factor
Right Tail						
no interaction	422.8	(1.3, -)	36	5.8	(1.3, -)	2.91
interaction	396.9	(2.08, -)	35	19.1	(1.3, -)	16.74
Left Tail						
no interaction	353.1	(-, .4)	36	39.6	(-, .07)	36.03
interaction	287.7	(-, .07)	35	71.3	(-, .07)	67.87
Both Tails						
no interaction	273.7	(.1, -.2)	35	74.6	(.2, -.1)	67.24
interaction	255.0	(.35, -.1)	34	83.6	(.35, -.1)	76.84