Czado, Munk:

Noncanonical Links in Generalized Linear Models - When is the Effort Justified?

Projektpartner

Noncanonical Links in Generalized Linear Models - When is the Effort Justified?

Claudia Czado[†] and Axel Munk [*]

Abstract

Generalized linear models (GLM) allow for a wide range of statistical models for regression data. In particular, the logistic model is usually applied for binomial observations. Canonical links for GLM's such as the logit link in the binomial case, are often used because in this case sufficient statistics for the regression parameter exist which allow for simple interpretation of the results. However, in some applications, the overall fit as measured by the p-values of goodness of fit statistics (as the residual deviance) can be improved significantly by the use of a noncanonical link. In this case, the interpretation of the influence of the covariables is more complicated compared to GLM's with canonical link functions. It will be illustrated through simulation that the p-value associated with the common goodness of link tests is not appropriate to quantify the changes to mean response estimates and other quantities of interest when switching to a noncanonical link. In particular, the rate of misspecifications becomes considerably large, when the inverse information value associated with the underlying parametric link model increases. This shows that the classical tests are often too sensitive, in particular, when the number of observations is large. The consideration of a generalized p-value function is proposed instead, which allows the exact quantification of a suitable distance to the canonical model at a controlled error rate. Corresponding tests for validating or discriminating the canonical model can easily performed by means of this function.

keywords: generalized linear models, goodness of link tests, logistic regression, link function, parametric links, model validation and discrimination, p-value curve.

1 Introduction

Generalized linear models (GLM) allow for the treatment of regression problems in which the response can be non normally distributed. More specifically, the response distribution can be any distribution in a one parameter exponential family. This includes normal, binomial, Poisson, gamma and inverse Gaussian responses (see McCullagh & Nelder (1989)). In addition, a link function connecting the mean response with the linear predictor has to be chosen. GLM's with canonical links, such as the logit link in binomial regression (for a definition of a canonical link see McCullagh & Nelder (1989)), guarantee maximum information and a simple interpretation of the regression parameters, because in this case we obtain a linear model for the natural parameter of the underlying exponential family. For example,

the logit link gives a simple representation of the odds, which aids the interpretation of the results. In addition, the concavity of the likelihood function guarantees uniqueness of the MLE.

Canonical links, however, do not always provide the best fit available to a given data set. In this case, the link could be misspecified, which can yield substantial bias for the regression parameter estimates as well as for the mean response estimates (for example see Czado and Santner (1992) in the binomial response case). The most common approach to guard against such a misspecification, is to embed the canonical link into a wide parametric class of links $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$, which includes the canonical link as a special case when $\psi = \psi_*$, say. Many such parametric classes of link functions for binary regression data have been proposed in the literature. For example, Van Montford 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. The general case of link functions in GLM's was studied by Pregibon (1980) and Czado (1992). Czado (1995) developed criteria on how to choose such a family.

It should be noted, that a link misspecification represents only a special systematic departure from the model, while misspecification of the variance function or scales of the covariates are other possible systematic departures from the model. The methods for checking such departures are similar to the ones developed for the link (see McCullagh and Nelder (1989), Chapter 12) and therefore we restrict attention in this paper to the problem of link validation.

In the following, it is assumed that the true underlying link is a member of such a class $\mathcal{F}$. To protect against link misspecification large sample tests such as the likelihood ratio and the score test are recommended (Pregibon (1980, 1982) and McCullagh and Nelder (1989), Chapter 11) to assess, if a different link will lead to a significant improvement in fit. Hence guarding against link misspecification becomes tantamount to the testing problem

$$H : \psi = \psi_* \qquad \text{versus} \qquad K : \psi \neq \psi_*. \tag{1}$$

If H cannot be rejected, the additional consideration of the p-value is accepted as a sufficient measure for the evidence to keep the canonical model. If H is rejected, maximum likelihood estimation (MLE) is required to estimate jointly the link parameter ψ and the regression parameters. Unfortunately, this increases significantly the computational effort to analyze the data, because a noncanonical link model $\mathcal{F}$ cannot be performed in standard software packages and special software has to be written. In addition to the special software requirement, the estimation of the link also inflates the variance of the regression parameters, since the link parameter ψ cannot be chosen orthogonal in the sense of Cox and Reid (1987) to the regression parameters (see Taylor (1988) and Czado (1995)). Further, checks for isolated departures from the model (for a general review see Davison and Tsai (1992)) have been developed so far only for fixed link models (see Pregibon (1981) for logistic regression, Lee (1987, 1988) and Williams (1987) for GLM's).

Therefore, the goal of this paper is to investigate and answer the following two questions:

Q1 When is the effort justified to switch from a canonical link to a noncanonical link in a GLM ? (Model Discrimination)

Q2 How large is the evidence for the canonical model indicated by a large p-value associated to a goodness of link test for (1)? (Model Validation)

Within the classical approach, both questions are assumed to be sufficiently answered by the consideration of the p-value associated to one of the above mentioned standard tests. It is the aim of this

paper to show in a first step, that this way of proceeding provides in general no information concerning both of these questions Q1, Q2. Even after diagnostic tools are used in order to protect against outliers or other isolated deviations from the model, this approach cannot be justified. In a second step, we suggest an alternative procedure, which allows inference about the precise deviation from the canonical model.

To fix ideas, we restrict for the moment our consideration to the large class of generalized logistic links introduced by Czado (1992), even though any other parametric class mentioned above could have been used. However, this class of Box-Cox transformations of the linear predictor is preferred because it allows separate modification of the right and left tail of the link function and its parametrization is locally orthogonal (Czado (1995)). Note, that for this family $\psi = \psi_* = 1$ always corresponds to the canonical link. Consider now the following binomial regression data sets:

Example 1: (Age of Menarche in Warsaw Girls)
Milicer and Szczotka (1966) analyzed the occurrence of menarche as a function of age in a sample of 3918 Warsaw girls (see Table 3 of Stukel (1988) for data). The standard analysis based on the logistic link with age as covariate reveals lack of fit, especially in the left tail. Table 1.1 summarizes the results giving parameter estimates and their estimated standard errors in parentheses in the first column. Residual deviances, their degrees of freedom and the p-value of corresponding goodness of fit test in parentheses are given in the second column. The likelihood ratio statistic for testing logistic link by (1), their degree of freedom and the corresponding p-value in parentheses is provided in the third column of the table.

Model	Estimated Link ψ	Residual Deviance	Likelihood Ratio
logistic		26.70 (23, .27)	
right tail	.88 (.083)	25.09 (22, .29)	1.61 (1, .204)
left tail	1.39 (.138)	17.62 (22, .73)	9.08 (1, .003)

Table 1.1: Link Estimates, Residual Deviance and Likelihood Ratio Statistics for the Age of Menarche Data

Therefore, following the usual way of proceeding we would decide for the noncanonical model with left tail modification. We will see in the following that this decision is only justified, when the odds is the parameter of interest.

Example 2: (Bottle Deposit Data)
Neter, Wasserman & Kutner, p. 617 (1989) gave data on the number of bottles returned for 6 different levels of deposits. The results of a generalized logistic analysis are contained in Table 1.2.

Model	Estimated Link ψ	Residual Deviance	Likelihood Ratio
logistic		12.18 (4, .02)	
right tail	1.56 (.231)	5.28 (3, .15)	6.90 (1, .009)
left tail	.63 (.197)	10.03 (3, .02)	2.15 (1, .143)

Table 1.2: Link Estimates, Residual Deviance and Likelihood Ratio Statistics for the Bottle Deposit Data

Here we are left in a somewhat ambiguous situation. Although the LR-test supports a right tail modification with high evidence (p-value = .009) the residual deviance only indicates a slight improvement

in fit (p-value $= .15$). Therefore, one would argue that a right tail modification is necessary because the (type I) error of this decision is controlled at a rather small rate.

Although a noncanonical model is always more flexible and improves the fit when used, there are still strong reasons why the experimenter wants to retain the canonical link. The cost of a noncanonical link (as described above) is not justified, for example, if the effects of using this link instead of the more "fitting" noncanonical link are small on the mean response estimates or other quantities of particular interest. We will show, that the effects on the mean response estimates are about the same for both data sets, although the effects are very different on the estimated odds. More specifically, the odds are changed up to a factor of 50 for the menarche data set, while they are changed only up to a factor of 1.5 for the deposit data. Therefore, if the odds is the parameter of interest, a noncanonical link is truly needed for the menarche data, while it is not necessary for the deposit data. One could also be interested in other quantities as the odds which affects the above conclusions. The choice of such alternative measures of discrepancy between the canonical and noncanonical model will be discussed carefully in Section 3.

We will make more precise in Section 4.1. by means of a simulation study what the above examples indicated — that the use of the classical tests may urge the experimenter to decide for the (much more complicated) noncanonical model (discrimination in the sense of Q1), although information for the parameter of interest is not (scientifically relevant) increased. In addition, it turns out that the problem of model validation (Q2) also cannot be answered sufficiently by using the above mentioned large sample point hypotheses tests. Roughly speaking, this study indicates two systematic errors. First, when the data does not provide a precise estimation of the link parameter, the classical tests will not reject $H : \psi = \psi^*$ with large probability although the true mean response (or other parameters of interest) is far apart from the mean response under the canonical link assumption. Contrary, when the variation of the estimated link is small or the sample size is too large, we find that these tests lead with high probability to a decision in favor of the noncanonical model - although the effort is not justified, i.e. the difference between the true mean response and that in the canonical model is negligible small. We mention, that these problems are not caused by the particular choice of the goodness of link *test* for the testing problem (1), rather this is intrinsically related to the misleading *hypotheses H* in (1) and its corresponding p-values. Whenever the null hypothesis is rejected, no information about the discrepancy to the noncanonical model is involved — whereas acceptance of H (or even a large p-value) does provide no evidence in favor of the canonical model.

To remedy the above described problems, we suggest in a first step to consider tests for hypotheses of the form

$$H : \psi \notin [\psi_* - \psi_l, \psi_* + \psi_u] \text{ versus } K : \psi \in [\psi_* - \psi_l, \psi_* + \psi_u] \tag{2}$$

for specified $0 < \psi_l, \psi_u$. If H is rejected at some level α, the canonical link is validated with controlled error probability α within a (ψ_l, ψ_u)-neighborhood. In Section 3, we discuss several criteria on how to choose these bounds. This allows to quantify the deviation from the canonical model in terms of those quantities the experimenter is particularly interested in (e.g. the odds in a binomial response model). Defining the problem of link validation in terms of a Neyman Pearson test problem is always encountered with uncertainty — whenever the hypothesis is not rejected. Therefore, following Fisher, we suggest in a second step the consideration of a generalized p-value curve associated to the testing problem (2). This approach is more adapted to a decision theoretic formulation of the problem of model assessment. In particular, the evidence of H and K can simply be graphically illustrated by these curves. This gives the applied working statistician a simple tool at hand to visualize rapidly the goodness of fit at controlled error rates — for both, validation and discrimination.

4

Our approach is based on the asymptotic distribution of the joint maximum likelihood estimator of link and regression parameters. To this end, we extend in Section 2 results by Fahrmeir and Kaufmann (1985) and apply these to the construction of tests and p-value curves for the problem (2). The simulation results in Section 4.2 of this test for validating and discriminating a logistic link show that the significance level α is maintained in small samples, which allow the proposed tests to be used for the analysis of a link in GLM at controlled error rate.

In Section 5, we return to the examples presented above and illustrate generalized p-value curves in action.

2 Generalized Linear Models with Parametric Link

2.1 Asymptotic Theory

Ordinary GLM's have been extended to allow for data selected link functions from a class of parametric functions. For the binomial response case, this is evidenced by the many parametric link families considered in the literature. In the context of other GLM's, this extension was first considered by Pregibon (1980) and investigated in more detail by Czado (1992, 1995).

The following model for regression data with response Y_i and explanatory variables $X_i = (x_{i1}, \cdots x_{ip})$ for $i = 1, \cdots, n$ will be used:

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

$$f_{y_i}(y_i, \theta_i, \phi) = \exp[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)] \qquad (3)$$

   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 $\eta_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ for $1 \leq i \leq n$ influence the response Y_i. Here $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)$ 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 $\Im = \{F(\cdot, \psi) : \psi \in \Psi\}$.

Attention is restricted to link families $\Im$ which contain only strictly monotone continuous functions $F(\cdot, \psi)$ indexed by a scalar link parameter ψ. It should be noted that in conventional GLM terminology the link g is equal to the inverse of F. An unknown scale parameter ϕ in (3) is estimated by an appropriate moment estimator involving the Pearson χ^2 Statistic.

In GLM's with parametric link as defined by (3), the regression parameter $\boldsymbol{\beta}$ and the link parameter ψ are jointly estimated by maximum likelihood. If the true link F is a member of the link family $\Im$, the joint MLE $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}, \hat{\psi})$ of $\boldsymbol{\delta} = (\boldsymbol{\beta}, \psi)$ will be shown to be strongly consistent and efficient under regularity conditions. This asymptotic normal distribution of the joint MLE $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}, \hat{\psi})$ of $\boldsymbol{\delta} = (\boldsymbol{\beta}, \psi)$ will then be used to construct a validation test H versus K as given by (2)

As for ordinary GLM's, one has the relationship $\mu_i = \frac{d}{d\theta} b(\theta)|_{\theta=\theta_i} = b'(\theta_i)$. The log likelihood $l(\delta)$ derived from model (3) can be written as:

$$l(\delta) = \sum_{i=1}^{n} [\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)] \text{ where } \mu_i = b'(\theta_i) \text{ and } \mu_i = F(\eta_i(\beta), \psi).$$

To derive the corresponding scores, note that $\mu_i = F(\eta_i(\beta), \psi)$ holds, which implies

$$\frac{\partial \mu_i}{\partial \psi} = F_{i2}, \frac{\partial \mu_i}{\partial \beta_j} = x_{ij} F_{i1} \text{ for } 0 \le j \le p, 1 \le i \le n, \tag{4}$$

where $F_{i1} = \frac{\partial}{\partial \eta} F(\eta, \psi)|_{\eta=\eta_i}, F_{i2} = \frac{\partial}{\partial \psi} F(\eta, \psi)|_{\eta=\eta_i}$ and $x_{i0} = 1$ for $1 \le i \le n$. Let $d_i = \frac{d\theta_i}{d\mu_i}$ and use (4) to express the scores as follows:

$$s_j(\delta) = \frac{\partial}{\partial \beta_j} l(\delta) = \sum_{i=1}^{n} \frac{d\theta_i}{d\mu_i} \frac{\partial \mu_i}{\partial \beta_j} [\frac{y_i - \mu_i}{a(\phi)}] = a(\phi)^{-1} \sum_{i=1}^{n} d_i x_{ij} F_{i1}(y_i - \mu_i), \tag{5}$$

$$s_{p+1}(\delta) = \frac{\partial}{\partial \psi} l(\delta) = \sum_{i=1}^{n} \frac{d\theta_i}{d\mu_i} \frac{\partial \mu_i}{\partial \psi} [\frac{y_i - \mu_i}{a(\phi)}] = a(\phi)^{-1} \sum_{i=1}^{n} d_i F_{i2}(y_i - \mu_i). \tag{6}$$

To compute the joint MLE $\hat{\delta}$, solve simultaneously the equations obtained by equating the scores (5) to zero.

Finally, the expected Fisher information $I_n(\delta)$ for model (3) can be expressed as follows:

$$I_n(\delta) = a(\phi)^{-1} \left[\begin{array}{cc} I_{\beta,\beta} & I'_{\beta,\psi} \\ I_{\beta,\psi} & I_{\psi,\psi} \end{array} \right], \tag{7}$$

where $I_{\beta,\beta}$ is a (p+1)x(p+1) matrix, $I_{\beta,\psi}$ is a (p+1) vector and $I_{\psi,\psi}$ is a real number given by

$$(I_{\beta,\beta})_{rs} = \sum_{i=1}^{n} x_{is} x_{ir} F_{i1}^2 d_i$$

$$(I_{\beta,\psi})_r = \sum_{i=1}^{n} x_{ir} F_{i1} F_{i2} d_i \ \ 0 \le r, s \le p \text{ and } I_{\psi,\psi} = \sum_{i=1}^{n} F_{i2}^2 d_i.$$

It is straight forward to see, that the score vector $s_n(\delta) = (s_1(\delta), \cdots, s_{p+1}(\delta))$ has covariance matrix $I_n(\delta)$. Let $H_n(\delta)$ denote the corresponding observed information matrix, that is the (s,t)th element of $H_n(\delta)$ is given by

$$H_n(\delta)_{st} = \frac{\partial^2 l(\delta)}{\partial \delta_s \delta_t} \text{ for } s, t = 1, \cdots, p+1.$$

The minimal (maximal) eigenvalue of a square matrix A will be written as $\lambda_{min}(A)(\lambda_{max}(A))$. Let $\delta_0 = (\beta_0, \psi_0)$ denote the true parameter values. For brevity, we will write I_n and s_n for $I_n(\delta_0)$ and $s_n(\delta_0)$, respectively. The following regularity conditions are needed:

R1 $\lambda_{min}(I_n) \to \infty$ as $n \to \infty$.

R2 There is a neighborhood $N \subset B$ of δ_0 such that a.s

$$\lambda_{min}(H_n(\delta)) \ge c[\lambda_{max} I_n]^{\frac{1}{2}+\varepsilon}, \delta \in N, n \ge n_1$$

with some constants $c, \varepsilon > 0$ and a random number n_1.

R3 Assume $\{\mathbf{x}_n, n \geq 1\} \subset K$ compact and $F(x_n^t \boldsymbol{\beta}, \psi)$ twice differentiable with respect to $\boldsymbol{\beta}$ and ψ and bounded for $\{\mathbf{x}_n, n \geq 1\} \subset K$ for fixed β and ψ,

R4 $\frac{n}{\lambda_{min}(I_n)}$ is uniformly bounded $\forall n \geq 1$.

The following results are straight forward modifications of results for ordinary links previously obtained by Fahrmeir & Kaufmann (1985). Observe, that additional estimation of the link parameter requires slightly stronger assumptions on the link function $F(\cdot, \psi)$. Here $F(\eta, \psi)$ has to be also twice differentiable with regard to the link parameter ψ.

Theorem 2.1 *Under (R1) und (R2) with $\varepsilon > 0$, there is a sequence $\hat{\boldsymbol{\delta}}_n$ of random variables and a random number n_1 with*

(i) $P(s_n(\hat{\boldsymbol{\delta}}_n) = \mathbf{0} \text{ for all } n \geq n_1) = 1$ *(asymptotic existence),*

(ii) $\hat{\boldsymbol{\delta}}_n \to \boldsymbol{\delta}_0$ *a.s. (strong consistency)*

We are now in the position to give the asymptotic result:

Theorem 2.2 *Under (R1), (R3) and (R4), there is a sequence $\hat{\boldsymbol{\delta}}_n$ such that $P(s_n(\hat{\boldsymbol{\delta}}_n) = \mathbf{0}) \to 1$ as $n \to \infty$ and*

$$I_n^{\frac{1}{2}}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \xrightarrow{\mathcal{D}} N_{p+1}(\mathbf{0}, I) \qquad \text{as } n \to \infty,$$

where $N_m(\mu, \Sigma)$ denotes an m-dimensional normal distribution with mean vector μ and variance-covariance matrix Σ.

Sketches of the proofs of these theorems are contained in the appendix. They are extensions of results of Fahrmeir and Kaufmann (1985). The binary case with parametric link has been previously considered by Czado (1989).

2.2 Asymptotic Link Validation Tests in Generalized Linear Models

Using the results of Section 2.1, we are now able to construct an asymptotic link validation test for GLM's.

Theorem 2.3 *Under the assumptions of Theorem 2.2 for the validation problem H versus K in (2) a consistent asymptotic level α test is given by the rejection region*

$$\mathcal{C} = \left\{ \hat{\psi}_n : \left| \Phi\left(\frac{\hat{\psi}_n - \psi_* - \psi_u}{\hat{\sigma}_n(\psi, \beta)} \right) - \Phi\left(\frac{-(\hat{\psi}_n - \psi_*) + \psi_l}{\hat{\sigma}_n(\psi, \beta)} \right) \right| \leq \alpha \right\} \tag{8}$$

where $\hat{\sigma}_n^2$ denotes any consistent estimate of $\sigma_n^2(\psi, \boldsymbol{\beta})$ which is the $(p+1, p+1)$-th element of the inverse of the Fisher information matrix $I_n(\boldsymbol{\delta})$.

The critical region of a test at level $1 - \alpha$ for the discrimination problem K versus H is given by the complement $\mathcal{C}^c$.

Proof. Fix $\hat{\sigma}^2$. $\mathcal{C}$ can be rewritten as

$$\mathcal{C} \quad := \quad \left\{ \hat{\psi}_n \ : \ \hat{\psi}_n \in [-C, C] \right\}$$

where $C > 0$ is uniquely determined by

$$\hat{P}(C, \hat{\sigma}_n, \psi_l, \psi_u) \quad := \quad \Phi\left(\frac{C - \psi_* - \psi_u}{\hat{\sigma}_n} \right) - \Phi\left(\frac{-(C + \psi_*) - \psi_l}{\hat{\sigma}_n} \right) = \alpha. \tag{9}$$

Here we have used that the family of normal distributions parametrized by the mean and fixed variance $\hat{\sigma}_n^2$ is a one-parametric exponential family and hence totally positive of order 3 (cf. Lehmann (1986), Th.6, p.101). Note further, that condition (25) of Lehmann (1986, p.102) reduces to (9) because of the symmetry of the distribution of $\hat{\psi}_n - \psi_* - (\psi_u + \psi_l)/2$. From Theorem 2.2 we draw that

$$\mathcal{L}\left\{ (\hat{\psi}_n - \psi)\sigma_n^{-1}(\beta, \psi) \right\} \quad \longrightarrow \quad \mathcal{N}(0, 1) \qquad \text{as } n \to \infty.$$

Applying Slutzky's Theorem proves that the test is asymptotic size α. Consistency is similar.

As an estimator of $\sigma_n^2(\psi, \beta)$ we can use $\hat{\sigma}_n^2 := \sigma_n^2(\hat{\psi}, \hat{\beta})$ where $(\hat{\psi}, \hat{\beta})$ denotes the joint MLE of (ψ, β).

The function $\hat{P}$ in (9) plays a central role for the assessment of a logistic model and will be denoted in the following as (asymptotic) generalized p-value function associated to the validation and discrimination problem (H versus K and converse) in (2). Observe, that given a fixed sample of observations $\hat{P}(\hat{\psi}_n, \hat{\sigma}_n, \psi_l, \psi_u)$ can be regarded as a two $-$ dimensional surface, where the level sets $\alpha = \hat{P}$ give the asymptotic minimal bounds ψ_l and ψ_u for which H can be rejected at level α as well as the maximal bounds for which K can be rejected by the discrimination test at level $1 - \alpha$. In particular, when $\psi_l = \psi_u = \psi_*$, 1-$\hat{P}$ denotes the 'classical' p-value of the maximum likelihood test for K against H. $\hat{P}$ and its complement can be (asymptotically) regarded as a *precise* measure of the evidence of neighborhoods (ψ_l, ψ_u) in contrast to the classical 'two-sided' p-value associated to (1). For an illustration of $\hat{P}$ we defer to the examples discussed in Section 5.

3 Determining the Tolerance Bound

Crucial for the specification of the hypotheses H and K in (2) are the values of the tolerance constants $\psi_l > 0$ and $\psi_u > 0$. We will discuss several choices depending on the quantities one is interested in estimating. For this let $\hat{\beta}(\psi)$ denote the MLE of β, when a fixed link parameter ψ of an arbitrary parametric link family $\mathcal{F} = \{F(\cdot, \psi), \psi \in \Psi\}$ is used and $\eta_i(\hat{\beta}(\psi))$ denotes the corresponding i-th linear predictor. As a reasonable measure of discrepancy from the canonical model the maximal change in the mean response estimates when switching from the canonical link ψ_* to the link ψ could be used. This can be estimated by

$$\Delta_m(\psi) = \max_{i=1,\cdots,n} |F(\eta_i(\hat{\beta}(\psi)), \psi) - F(\eta_i(\hat{\beta}(\psi_*)), \psi_*)|.$$

Note, that in the case of binomial responses Δ_m weights differences in the tail the same as differences in the middle of the success probabilities. Often, however, we are only interested in the relative (and not absolute) changes to the mean response estimates. Therefore, consideration of the maximal change in the odds might be useful when the experimenter has difficulties assessing the importance of changes in mean responses. This is estimated by

$$\Delta_o(\psi) = \max_{i=1,\cdots,n} \left\{ \begin{array}{ll} o_i(\psi) & \text{if } F(\eta_i(\hat{\beta}(\psi)), \psi) > F(\eta_i(\hat{\beta}(\psi_*)), \psi_*) \\ \frac{1}{o_i(\psi)} & \text{otherwise} \end{array} \right. , \tag{10}$$

where

$$o_i(\psi) = \left[\frac{F(\eta_i(\hat{\beta}(\psi)), \psi)}{1 - F(\eta_i(\hat{\beta}(\psi)), \psi)} \right] \bigg/ \left[\frac{F(\eta_i(\hat{\beta}(\psi_*)), \psi_*)}{1 - F(\eta_i(\hat{\beta}(\psi_*)), \psi_*)} \right].$$

Note that deviations from the logistic odds are measured from both directions in (10).

Particularly, in the binomial case we can even compare the maximal absolute difference between all possible success probabilities under a logistic model and a model with link parameter ψ, given by

$$\Delta_p(\psi) = sup_{\eta \in \mathcal{R}} |F(\eta, \psi) - F(\eta, \psi_*)|. \tag{11}$$

Note that, for binomial responses the supremum in (11) is finite. This criterion might be particularly appealing for the applied statistician because it has not to be estimated as Δ_o and Δ_m.

Once decided for a criterion as $\Delta_m(\psi)$, $\Delta_o(\psi)$ or $\Delta_p(\psi)$, we can determine the corresponding tolerance bound Δ_0, for which the canonical model is assumed as sufficiently approximated by the noncanonical model. Testing $\tilde{H} : \Delta_{(\cdot)}(\psi) > \Delta_0$ (or $\tilde{K} : \Delta_{(\cdot)}(\psi) \leq \Delta_0$) is now tantamount to the testing problem (2) and its converse, where the constants ψ_l, ψ_u have to be determined numerically as $\Delta_{(\cdot)}^{-1}(\Delta_0)$. Observe, that this leads always to two unique values $\psi_l^0 < \psi_u^0$, such that $\psi_l^0 < \psi_* < \psi_u^0$ because each criterion $\Delta_{(\cdot)}(\psi)$ is a strictly unimodal function with unique minimum at ψ_*. The determination of the bounds ψ_l, ψ_u can now be expressed in terms of one of the suggested measures of discrepancy $\Delta_{(\cdot)}$ in accordance with the particular model and the specific question the experimenter has in mind. This allows to reduce the two dimensional surface $\hat{P}$ to a one - dimensional curve which only depends on $\Delta_{(\cdot)}$ by the relation $\hat{P}(\cdot, \cdot, \Delta_0) := \hat{P}(\cdot, \cdot, \psi_l^0, \psi_u^0)$. For illustration in the case of generalized logistic regression we displayed some values for the criterion Δ_p in Table 4.1.

4 Simulation Results

4.1 Small Sample Behavior of Goodness of Fit Point Hypothesis Tests in a Logistic Model

To investigate the small sample behavior of the residual deviance as goodness of fit statistics and the likelihood ratio statistics for testing a logistic link, we consider a family of generalized logistic distributions $\{F(\cdot, \psi), \psi \in \Re\}$ (Czado (1992)) with heavier right tail ($\psi < 1$) and lighter right tail ($\psi < 1$) than the logistic distribution ($\psi = 1$). These links have low variance inflation (Taylor (1988)) due to the fact that the parametrization is orthogonal in a neighborhood around $\beta = 0$. In addition, they are location and scale invariant (cf. Czado (1995)). In particular, the distribution functions $F(\cdot, \psi)$ are given by

$$F(\eta, \psi) = \frac{\exp(h(\eta, \psi))}{1 + \exp(h(\eta, \psi))},$$

where

$$h(\eta, \psi) = \begin{cases} \frac{(\eta+1)^\psi - 1}{\psi} & \text{if } \eta > 0 \\ \eta & \text{otherwise} \end{cases}.$$

A modification of the left tail or both tails is also possible.

Figure 4.1 gives the absolute difference in probability between the logistic and the generalized logistic distribution as a function of η and ψ. It can be seen that for $\psi > 1$ (lighter right tail) this difference

9

is significantly large in a much smaller range of η values compared to the case of $\psi < 1$ (heavier right tail). This allows us to classify four areas of varying degree of information about the link parameter ψ. In the case of a heavier right tail ($\psi < 1$) compared to the logistic link ($\psi = 1$) *and* a large range for the linear predictors η_i it will be easy to discriminate against the logistic link, while the opposite will be true in the case when there is small range for the linear predictors. For the lighter right tail case ($\psi > 1$), the degree of information for discriminating against the logistic link will be medium in both cases of a large or small range for the linear predictors η_i.

The maximal distance $\Delta_p(\psi)$ between the generalized logistic link and the logistic link as a function of the link parameter ψ is given in Figure 4.2. Table 4.1 gives for some special Δ_0 values the corresponding interval $[\psi_l, \psi_u]$ to insure that $\Delta_p(\psi) < \Delta_0$ for all $\psi \in [\psi_l, \psi_u]$.

Δ_0	ψ_l	ψ_u
.2	-.28	3.10
.15	.02	2.40
.10	.35	1.84
.075	.49	1.60
.05	.65	1.39
.025	.82	1.19
.01	.93	1.07

Table 4.1: Choice of the link parameter ψ to achieve a maximal absolute difference of Δ_0 between the generalized logistic cdf and logistic cdf

Data from the following generalized logistic model with single covariate was generated:

$$Y_i \simeq \text{binomial}(n, p_i) \text{ for } i = 1, \cdots, 11 \text{ with } p_i = F(\beta_0 + \beta_1 x_i, \psi),$$

where $F(\eta, \psi)$ is given by (4.1) and $\beta_0 = 1$. To investigate equally and unequally spaced covariate values two choices for $\boldsymbol{x} = (x_1, \cdots, x_{11})$ were considered, $\boldsymbol{x}$ equally spaced between -5 and 5 and $\boldsymbol{x}$ a standard normal random sample of size 11. The following values for the regression slope parameter β_1 were chosen: $\beta_1 = .5, 1, 2$. This allows models with nearly symmetric true probabilities around .5 ($\beta_1 = 1$), extreme probabilities ($\beta_1 = 2$) and more central probabilities ($\beta_1 = .5$). Figure 4.3 plots the true probabilities when $\psi = 0.02, 1, 2.4$ for the different β_1 and x values. Note that the equally spaced covariate case will induce a large range for the linear predictors, while the unequally spaced covariate case will induce a small one. Therefore, we suggest to classify the cases $\psi < 1$ and equally spaced covariates ($\psi < 1$ and unequally spaced covariates) as areas with high (low) power to discriminate against the logistic link. The other areas ($\psi > 1$ and both cases of covariate configuration) have medium discrimination power. This will be supported by the following simulation results. Finally, we investigated two binomial sample sizes of $n = 20$ and $n = 40$.

To demonstrate the inappropriateness of using a large p-value of the ordinary goodness of fit statistics as an indicator of a good fitting model, we simultaneously calculated the p-values of the likelihood ratio test of testing $\psi = 1$ as well as the residual deviance test assuming a logistic model based on 500 replications. Values for the link parameter ψ were chosen between .02 and 2.4 to allow up to 15% percent of absolute difference in the probabilities between the logistic and generalized logistic model. It has been checked, that this maximal difference does occur in the range of true probabilities assumed for the simulation design. We recorded the percentage of cases, where the p-value of the likelihood ratio test statistic was larger than .2 and .5, respectively (see Figure 4.4 and 4.5) and the percentage

10

where the p-value of the residual deviance statistic was larger than .5 and .75, respectively (see Figure 4.6 and 4.7) based on 500 replications. Different p-value bounds for the likelihood ratio statistic and the residual deviance were chosen, since the likelihood ratio test is primarily used as a test to detect deviation from the canonical link, while the residual deviance test is used as a goodness of fit test where it is common practice to assume a higher p-value as indication of a good fitting model.

We will now discuss the conclusions from the above plots according to the degree of information available for the discrimination from the logistic link. Recall that $\psi < 1$ and equally spaced covariates contain high, $\psi < 1$ and unequally spaced covariates contain low and $\psi > 1$ contain medium information about the link.

If one considers (as it is common practice) a a p-value larger than .2 for the likelihood ratio test as indication of a good fitting model, the test will be unable to detect the large maximal difference of 15% (10%) in probability up to 12.0% (6.4%) in the area of high, up to 28.2% (27.7%) in the area of medium and up to 46.7% (44.8%) in the area of low discrimination power when $n = 20$ ($n = 40$). The percentages are roughly halved when the proportion of cases for the likelihood ratio test statistics with a p-value greater than .5 are considered. Here, up to 3.6% (2.0%) in the area of high, up to 12% (10.2%) in the area of medium and 23.1%(23.4%) in the area of low discrimination information of the tests are unable to detect a difference in 15% (10%) probabilities when $n = 20 (n = 40)$. This shows that the likelihood ratio test is performing very poorly in areas of medium and low discrimination power, since it accepts the canonical link a large number of times, when in fact the true probabilities differ substantially like by 10-15% from the canonical link probabilities.

We now turn to the performance of the residual deviance test as a goodness of fit test. If one relies only on a residual deviance goodness of fit statistic as measure of goodness assuming a p-value of larger than .5, say, as an indication of a good fitting model, one can see that this test is especially unable to detect link misspecification when $\psi > 1$, i.e. in the area of medium discrimination power. In particular, we observe that up to 8.4% (7.8%) in the area of high, up to 64.9% (44.7%) in the area of medium and 31.4% (36.2%) in the area of low information of the residual deviance test are unable to detect a maximal absolute difference of 15% (10%) in probabilities when $n = 20$ ($n = 40$) assuming a p-value of .5 as indication of a good fitting model. Again, these percentages are roughly halved when a p-value of .75 is assumed as sufficient evidence for the canonical model.

In a second step, we determined the sensitivity of these two tests, i.e we are interested in the number of times the test would reject the canonical model, when in reality there is at most a negligible deviation from the canonical model. For this, we assumed a maximal absolute difference of 5% in probabilities as a negligible deviation from the logistic model. It turns out, that the residual deviance test has less sensitivity against small deviations from the canonical model than the likelihood ratio test. Both tests, however, are too sensitive in areas of high discrimination power and when the sample size is large ($n = 40$). In particular for the likelihood ratio test, we observe that for $n = 20$ ($n = 40$) up to 20.8% (34%) in the area of high, up to 15% (23.2%) in the area of medium and up to 9.2% (13%) in the area of low discrimination power to reject the logistic model at $\alpha = .05$ when the true underlying model only deviates by at most 5% in the probabilities from the logistic model. For the residual deviance test, the same percentages are 11.4% (28.2%) in the area of high, 4.8% (7%) in the area of medium and 8.6% (8.2%) in the area of low discrimination power when $n = 20$ ($n = 40$).

To summarize, these results clearly demonstrate, that on the one hand there is no guide on how large a p-value has to be, before it gives sufficient indication for a good fitting model. In any case, they have to be much larger than significance levels for rejecting the point null hypothesis. In particular, the residual deviance test turns out to be very poor in detecting a large deviation from the canonical model. In addition, prediction for some covariate values within the range of observed covariate values

11

will be completely unreliable. Therefore, ordinary goodness of fit tests such as the residual deviance test or even the likelihood ratio test for testing logistic link within the class of generalized logistic links should only be used very carefully to validate the logistic regression model. Large p-values turn out to be misleading. On the other hand, for certain (unknown) parameter configurations which provide high information about the link, both tests are too sensitive to the occurrence of deviations from the logistic model, which are too small to be of importance to the data analyst.

The Pearson χ^2 statistic assuming a logistic model has been also investigated. We obtained similar results for other GLM's which are not displayed by the ease of brevity.

4.2 Small Sample Properties of the Validation Test when Verifying the Logistic Model

We now investigate the small sample behavior of the proposed validation test for the special case of binomial responses. The same simulation setup as in the previous section has been used. Consistent with the previous section, we consider maximal absolute differences of 15% in probabilities from the logistic model as large deviation, while a maximal absolute difference of 5-10% are tolerable deviations from the link. Therefore, we investigated the performance of the asymptotic link validation test, when we allow a maximum absolute difference of 10% (5%) in probabilities corresponding to $\Delta_0 = .1$ ($\Delta_0 = .05$). This corresponds to the validation testing problem $H : \psi \notin [.35, 1.84]$ versus $K : \psi \in [.35, 1.84]$ ($H : \psi \notin [.65, 1.39]$ versus $K : \psi \in [.65, 1.39]$) when $\Delta_0 = .1$ ($\Delta_0 = .05$) . Here, $\psi_* = 1$ is equivalent to the logistic link. From the findings in the last section, we expect the power of the asymptotic link validation test to be larger for the equally spaced covariate case compared to the unequally spaced case and when the true link has a heavier right tail ($\psi < 1$) compared to a lighter right tail ($\psi > 1$).

Again two binomial sample sizes of n=20 (solid line) and n=40 (dotted line) were considered. Asymptotic validation tests of $H : \psi \notin [.35, 1.84]$ versus $K : \psi \in [.35, 1.84]$ ($H : \psi \notin [.65, 1.39]$ versus $K : \psi \in [.65, 1.39]$) when $\Delta_0 = .1$ ($\Delta_0 = .05$) at level $\alpha = .1$ were performed on 500 binomial data sets with $\psi = .02, .35, .65, 1, 1.39, 1.84, 2.4$ ($\psi = .35, .65, .82, 1, 1.19, 1.39, 1.84$) and the observed power is given in Figure 4.8 (Figure 4.9) for all cases considered. Cases where the maximization routine failed to converge were deleted from the simulation. This occurred in about 1% of all cases when $\psi = .35, .65, 1, 1.39$ and up to 7% for $\psi = .02, 1.84, 2.4$.

In all cases considered, the validation test allowing for 10% maximum absolute difference in probabilities maintains its significance level α of .1, with being more conservative on the left hand side of the alternative K ($\psi = .35$) and more liberal on the right hand side ($\psi = 1.84$). A possible explanation for this is the smaller area of large difference between the link parameters for $\psi > 1$ compared to $\psi < 1$. For the same reason, the power of the test is higher by about 50% for the equally spaced covariate case. The power of the test increases by about 50% as sample size changes from $n = 20$ to $n = 40$.

As to be expected, the validation test allowing only for a maximum absolute difference of 5% in probabilities has lower observed power than the test allowing for a maximum absolute difference of 10% in probabilities (see Figure 4.9). The reduction in power is large especially when $n = 20$ and for unequally spaced covariates. Even for a sample size of $n = 40$ the maximal power is .3, indicating that larger sample sizes than 40 are required. However, the test maintains equally well its significance level of $\alpha = .1$ at both end points of the alternative ($\psi = .65$ and $\psi = 1.39$).

In summary, the validation test maintains its significance level. The power of the test depends on whether data is collected in areas of large difference between the logistic link and the generalized logistic link. For a validation neighborhood of 10% in probabilities a sample size of $n = 20$ is sufficient for a maximal power of .5 in the case when the data can determine the areas of the large difference while

a sample size of $n = 40$ is needed for data which is sparse in areas of large difference. For the smaller validation neighborhood of 5% in probabilities, sample sizes larger than $n = 40$ are required. Hence, the proposed classification of regions of the parameter space into different zones of discrimination power is an excellent indicator for the actual power of the validation and discrimination test.

5 Examples

5.1 Age of Menarche in Warsaw Girls (Revisited)

For this data set, changes to the estimated success probabilities (see (11)) as well as to the estimated odds (see (10)) have been investigated. First, for a range of Δ_0 values the corresponding ψ_l and ψ_u values have been determined for both criteria. The corresponding generalized p-value functions (as defined in Section 2.2) are given in Figure 5.1. These functions were calculated as functions of the particular criterion $\Delta_{(\cdot)}$.

They show, that using a left tail modification of the link will result in a maximal absolute difference of 3% in estimated success probabilities at $\alpha = .1$ compared to a logistic analysis. A logistic analysis can be validated at $\alpha = .1$, if one is willing to tolerate a change of 7% in estimated probabilities. Since this data set contains extreme observed probabilities, it will be expected that the effects on the estimated maximal odds will be large, which is supported by Figure 5.1. In particular, a maximal change of the estimated odds by 5 can be detected, but the logistic link can only be validated when accepting maximal change by 50 on the estimated odds. Given these results, it seems to be reasonable that a noncanonical link model is necessary if the interest is focused on the odds. This is in accordance to the analysis made by the standard tests in the introduction. However, if we are only interested in the maximal probability difference, a modification of the model seems to be unnecessary, because a maximal probability difference under a left tail modification of ≤ 0.07 and ≤ 0.03 using a right tail modification can be validated at .1.

5.2 Bottle Deposit Data (Revisited)

Remember, that the standard likelihood ratio test gives strong indication for a noncanonical link model. As for the menarche data changes to the estimated mean responses as well as to the estimated odds are considered and the results are plotted in Figure 5.2. It shows, that a right tail modification will result in a maximal difference of 3% in estimated success probabilities compared to a logistic link analysis. This analysis can be validated at $\alpha = .1$, if one tolerates a change of 10% in probabilities. The maximal change on estimated odds is much less compared to the menarche data set. Here the logistic link can be validated in the neighborhood of a maximal change on the estimated odds of 1.48 at $\alpha = .1$. If the emphasis is on estimating odds, this change is certainly tolerable. Therefore, if the parameter of interest is the odds, it is certainly justified to maintain a logistic link despite the observed significant improvement in fit by the classical goodness of fit statistics, when a right tail modification is used. This example illustrates the too large sensitivity of the classical tests, when the sample size is large and the estimated variability in the link is small.

5.3 Toxicity of Insecticides to Flour Beetles

The following example shows that care has to be taken in the presence of isolated departures from the canonical model.

Collett (1991, p. 103) reports the results of an insecticidal trial on flour beetles which were sprayed in batches of about 50 insects with three different insecticides (DDT, γ-BHC, and DDT + γ-BHC) at varying deposits of spray. An initial analysis shows that a logistic model which fits parallel lines to the three different insecticides can be used as a basis for comparing the insecticides. Collett (1991, p. 142) gave arguments that this data set contains an outlier.

Right and left tail modifications to the link were investigated and link parameter estimates with their estimated standard error in parentheses, the residual deviance and the likelihood ratio statistics are reported in Table 5.1. for the complete data and the data with the outlier removed. This shows that the inflated residual deviance for the logistic model is entirely due to the presence of the outlier. However, in both cases there is no significant evidence that the link is misspecified. This shows that residual deviance statistics is very sensitive to isolated departures from the model, while the likelihood ratio statistics for testing logistic link is not in this example.

Model	Complete Data			Outlier removed		
	Link Estimate	Residual Deviance	Likelihood Ratio	Estimate Estimate	Residual Deviance	Likelihood Ratio
logistic		21.28 (14, .08)			14.84 (13, .32)	
right tail	1.20 (.25)	21.45 (13, .06)	.68 (1, .41)	1.12 (.23)	14.55 (12, .27)	.29 (1, .59)
left tail	1.43 (.37)	21.05 (13, .07)	1.1 (1, .30)	1.14 (.30)	14.65 (12, .26)	.19 (1, .66)

Table 5.1: Link Estimates, Residual Deviances and Likelihood Ratio Statistics for the Flour Beetle Data

In Figure 5.3 the p-value curves for assessing the maximal change to the estimated success probabilities for the complete data and the data with the outlier removed are given. No significant change compared to a logistic analysis can be shown for either data. The logistic link can be validated at $\alpha = .1$ in a neighborhood of a 6% maximal change to the success probabilities when the outlier is removed. This relatively large neighborhood can be explained by the only medium size of the binomial trials n_i. Since the observed probability of the outlier is 19/47 which is less than .5, the effect can be seen in the left tail of the link distribution. Here, the link can only be validated in a neighborhood of 10.5% maximal change to the probabilities when the outlier is left in the data. While the likelihood ratio statistics gives the impression that the logistic link is a perfect link for this data set, the asymptotic link validation test shows some uncertainty with regard to validating the logistic link due to the sparse information about the link contained in the data.

The changes to the estimated odds are given in Figure 5.4. Again the magnitude of the changes are reduced after the outlier is removed. Since more extreme probabilities are observed in the right tail than in the left tail, the changes to the odds are larger for a right tail modification than for a left tail one.

These last results show that isolated departures from the model such as the presence of outliers do certainly influence the performance the link validation test. It is to be expected that missing covariates and overdispersion in the data also influence the validation test. Therefore the link validation test should only be applied after diagnostic tools for the detection of a mean misspecification such as developed by Landwehr, Pregibon and Shoemaker (1984), Pregibon (1981), Williams (1987) and O'Hara Hines and Carter (1993) have been used. For the presence of overdispersion, score point hypothesis tests developed by Dean (1992) and Smith and Heitjan (1993) can be applied. Appropriate interval hypothesis tests, however, would be preferable over these score tests in the case of failing to reject the hypothesis of no overdispersion. The development of such tests is the subject of current research.

References

Aranda-Ordaz, F.J. (1981) On Two Families of Transformations to Additivity for Binary Response Data, *Biometrika*,**68**, 357-364.

Atkinson, A.C. (1982) Regression diagnostics, transformations and constructed variables (with discussion). *J. Roy. Statist. Soc. Ser. B*,**44**, 1-36.

Collett, D. (1991). *Modelling binary data*, Chapman and Hall, London.

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. *J. Roy. Statist. Soc. (B)*, **49**, 1-39.

Czado, C. (1989). Link Misspecification and Data Selected Transformations in Binary Regression Models, Ph.D. Thesis, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, N.Y., U.S.A.

Czado, C. and T.J. Santner (1992). The effect of link misspecification on binary regression inference. *J. Statist. Plan. Inf.***33**, 213-231.

Czado, C. (1992). On Link Selection in Generalized Linear Models. in *Advances in GLIM and Statistical Modelling, Lecture Notes in Statistics*, **78**, Springer Verlag, New York.

Czado, C. (1995). On Selecting Parametric Link Transformation Families in Generalized Linear Models. to appear in *J. Statist. Plan. Inf.*

Davison, A.C. and Tsai, C.-L. (1992). Regression Model Diagnostics, *Int. Statist. Rev.* , **60**, 3, 337-353.

Dean, C.B. (1992) Testing for overdispersion in Poisson and binomial regression models. *J. Amer. Stat. Assoc.*, **87**, 451-457.

Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13**, 342-368 (Correction: *Ann. Statist.* **14**, 1643).

Landwehr, J.M., Pregibon, D. and Shoemaker, A.C. (1984). Graphical methods for assessing logistic regression models. *J. Amer. Stat. Assoc.* , **79**, 61-83.

Lee, A.H. (1987). Diagnostic Displays for assessing Leverage and Influence in Generalized Linear Models. *Austral. J. Statist.*, **29**, 233-243.

Lee, A.H. (1988). Assessing Partial Influence in Generalized Linear Models, *Biometrics*, **44**, 71-77.

Lehmann, E.L. (1987). *Testing Statistical Hypotheses.* 2nd ed. New York, John Wiley.

O'Hara Hines, R.J. and Carter, E.M. (1993). Improved added variable and partial residual plots for the detection of influential observations in generalized linear models. *Appl. Statistics*, **42**, 3-20.

Guerrero, V.M. and Johnson, R.A. (1982). Use of the Box-Cox Transformation with Binary Response Models, *Biometrika*, **69**, 309-314.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition, Chapman and Hall, London.

Milicer, H. and Szczotka, F (1966). Age at Menarche in Warsaw Girls in 1965, *Human Biology*, **38**, 199-203.

Morgan, B.J.T. (1983). Observations on Quantit Analysis, *Biometrics*, **39**, 879-886.

Neter, J., Wasserman, W., Kutner, M.H. (1989) *Applied Linear Regression Models*, Second Edition, Irwin, Boston.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *J. Roy. Statist. Soc. Ser. C* **29**, 15-24.

Pregibon, D. (1981) Logistic regression diagnostics. *Ann. Statist*, **9**, 705-724.

Pregibon, D. (1982) Score Tests in GLIM with Applications. In GLIM82: *Proceedings of the International Conference on Generalized Linear Models, Lecture Notes in Statistics*, **13**. New York, Springer Verlag.

Prentice, R.L. (1976) A Generalization of the Probit and Logit Methods for Dose Response Curves, *Biometrics*, **32**, 761-768.

Smith, P.J. and Heitjan, D.F. (1993). Testing and adjusting for departures from nominal dispersion in generalized linear models. *Appl. Statistics*, . **42, 31-41**.

Stukel,T. (1988) Generalized logistic models. *J. Amer. Statist. Assoc.*, **83**, 426-431.

Taylor, J.M.G. (1988). The cost of generalized logistic regression. *J. Amer. Statist. Assoc.*, **83**, 1078-1083.

Van Montford, M.A.J. and Otten, A. (1976). Quantal Response Analysis: Enlargement of the logistic Model with a Kurtosis Parameter, *Biometrische Zeitschrift*, **18**, 371-380.

Whittmore, A.S. (1983). Transformations to Linearity in Binary Regression, *SIAM Journal on Applied Mathematics*, **43**, 703-710.

Williams, D.A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Appl. Statistics*, **36**, 181-191.

Appendix

Fahrmeir and Kaufmann (1985) proved the asymptotic results for ordinary GLM's, i.e for GLM's with fixed link. Their results need now to be extended to cover the case where a link parameter ψ has to be estimated as well.

Sketch of proof for Theorem 2.1 :
The proof of Theorem 2 (p. 349) of Fahrmeir and Kaufmann (1985) for noncanonical links (see Section 4.1) can be followed using a Taylor expansion of the log likelihood $l(\boldsymbol{\delta})$ around $\boldsymbol{\delta}_0$.

Sketch of proof for Theorem 2.2:
First, an analogue of Lemma 2 (Fahrmeir and Kaufmann (1985), p. 361) will be derived:

Lemma 5.1 *Under (R1) and (R3), $I_n s_n \xrightarrow{\mathcal{D}} N_{p+1}(\mathbf{0}, I)$ as $n \to \infty$.*

Proof of Lemma 5.1:
As in Fahrmeir and Kaufmann (1985), the proof uses the central limit theorem for triangular arrays and establishes the validity of the Lindeberg condition. For this, define the triangular array

$$Z_{ni} = \lambda^t I_n^{-\frac{1}{2}} s(y_i, x_i, \boldsymbol{\delta}_0)$$

where $s(y_i, x_i, \boldsymbol{\delta}_0)$ is the vector of individual score contributions, i.e. given by:

$$s(y_i, x_i, \boldsymbol{\delta}_0) = a(\phi)^{-1}(d_i x_{i1} F_{i1}(y_i - \mu_i), \cdots, d_i x_{ip} F_{i1}(y_i - \mu_i), d_i F_{i2}(y_i - \mu_i))$$

Define $\alpha_{ni} = \lambda^t I_n^{-\frac{1}{2}} \mathbf{L}(x_i^t \boldsymbol{\beta}_0, \psi_0)$, where

$$\mathbf{L}(x_i^t \boldsymbol{\beta}_0, \psi_0) = a(\phi)^{-1}(d_i x_{i1} F_{i1}, \cdots, d_i x_{ip} F_{i1}, d_i F_{i2}).$$

Note that this vector above is bounded when $\{x_n, n \geq 1\}$ by condition (R3), since d_i is a continuous function of $F(x_i^t \boldsymbol{\beta}_0, \psi_0)$. We can now express Z_{ni} as

$$Z_{ni} = \alpha_{ni}(Y_i - F(x_i^t \boldsymbol{\beta}_0, \psi_0)).$$

Under (R1) and (R3), we have with the Cauchy-Schwarz inequality

$$max_{i \leq n} \alpha_{ni}^2 \leq ||\mathbf{L}(x_i^t \boldsymbol{\beta}_0, \psi_0)||^2 \lambda_{min} I_n^{-1} \leq k \lambda_{min} I_n^{-1} \to 0 \text{ as } n \to \infty.$$

Then we argue as in Fahrmeir and Kaufmann (1985,p363) for compact regressors that the Lindeberg condition is satisfied.

We continue now with the sketch of proof for Theorem 2.1:

Using Lemma 3.1 and the following condition

R5 For every $\varepsilon > 0$

$$max_{\delta \in N_n(\varepsilon)} ||V_n(\boldsymbol{\delta}) - I|| \to 0 \text{ in probability}$$

where $V_n(\boldsymbol{\delta}) = I_n^{-\frac{1}{2}} H_n(\boldsymbol{\delta}) I_n^{-\frac{t}{2}}$ and $N_n(\varepsilon)$ is defined as $N_n(\varepsilon) = \{\boldsymbol{\delta} : ||I_n^{\frac{t}{2}}(\boldsymbol{\delta} - \boldsymbol{\delta}_0)|| \leq \varepsilon\}$,

we proceed as for the proof of Theorem 3 in Fahrmeir and Kaufmann (1985). Finally, it remains to show that (R1),(R3) and (R4) are sufficient for (R5). For this, we argue as in the proof of Theorem 4 (Fahrmeir and Kaufmann, p.364) considering the same partition of matrices as for $I_n(\boldsymbol{\delta})$ to adjust for the additional estimation of the link parameter ψ.
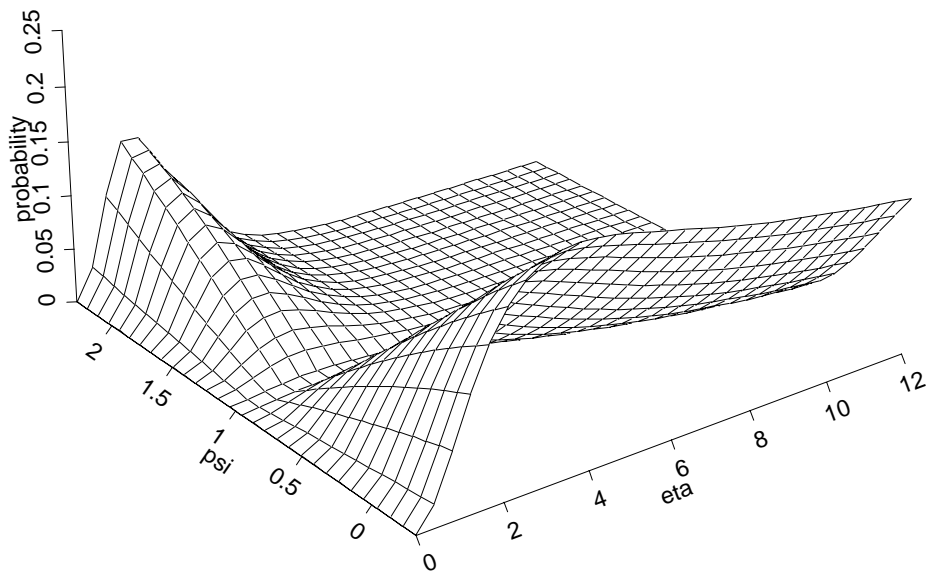
Figure 4.1: Absolute difference in probabilities between the logistic and generalized logistic distribution
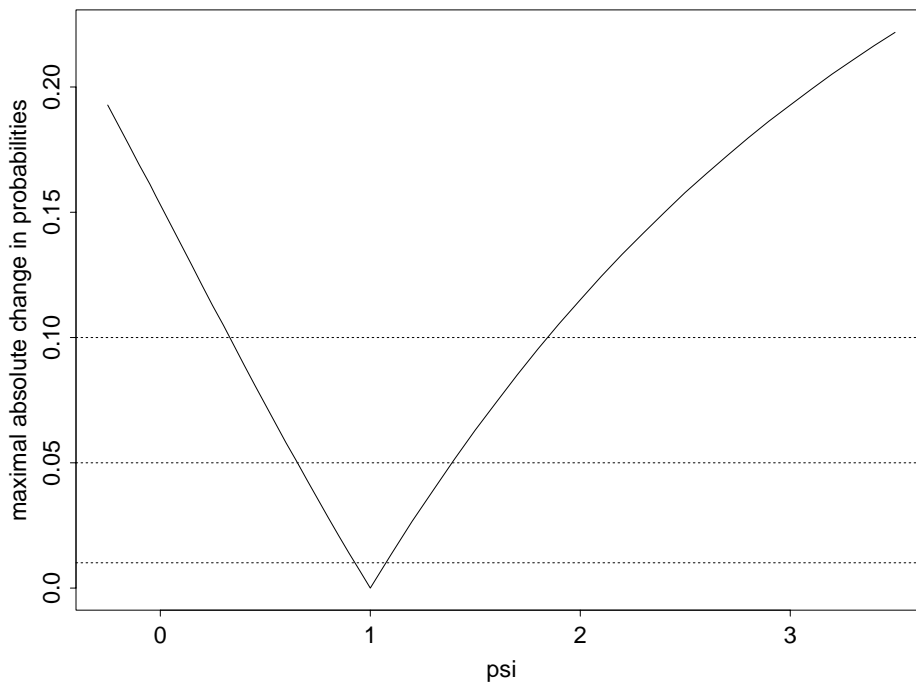
Figure 4.2: Maximal absolute difference $(\Delta_p(\psi))$ between the logistic and the generalized logistic distribution
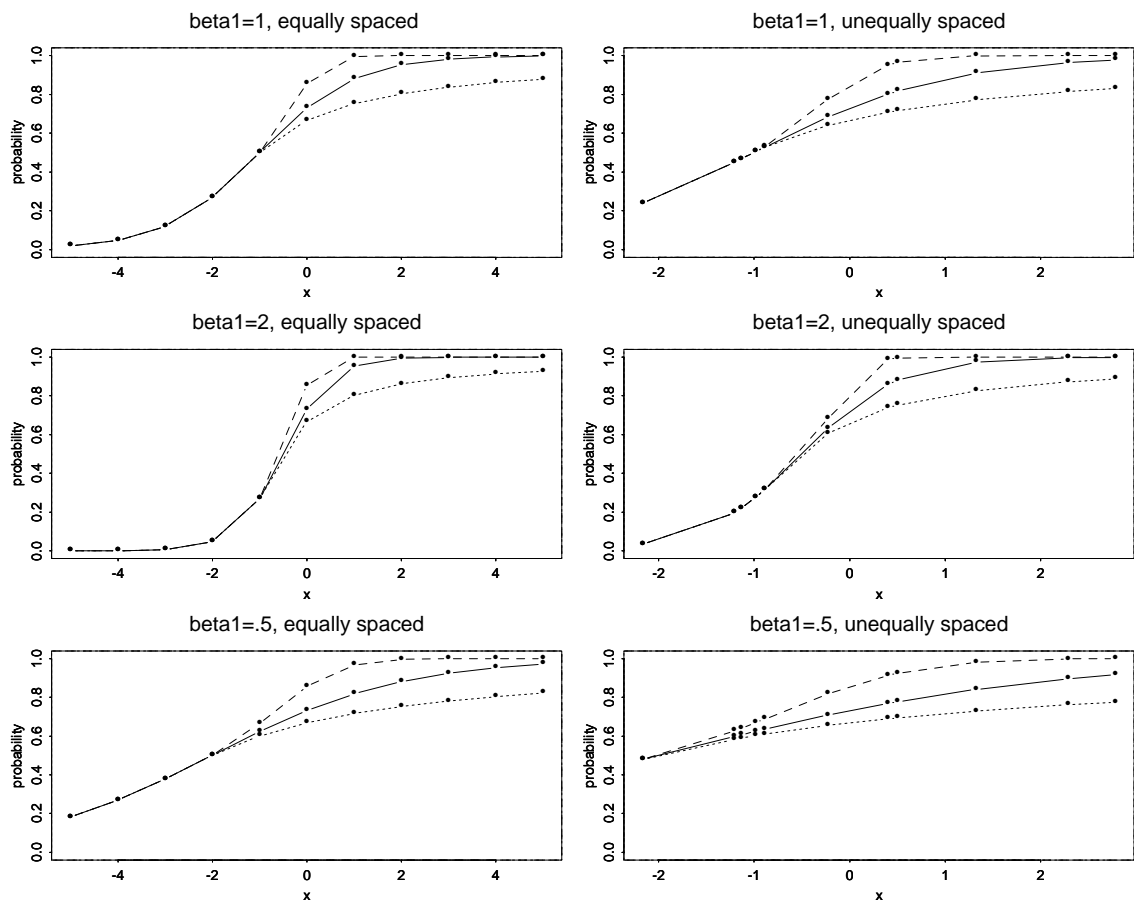
Figure 4.3: True link functions used in the simulation ($\psi = 1$ (solid line), $\psi = .02$ (dotted line), $\psi = 2.4$ (broken line))
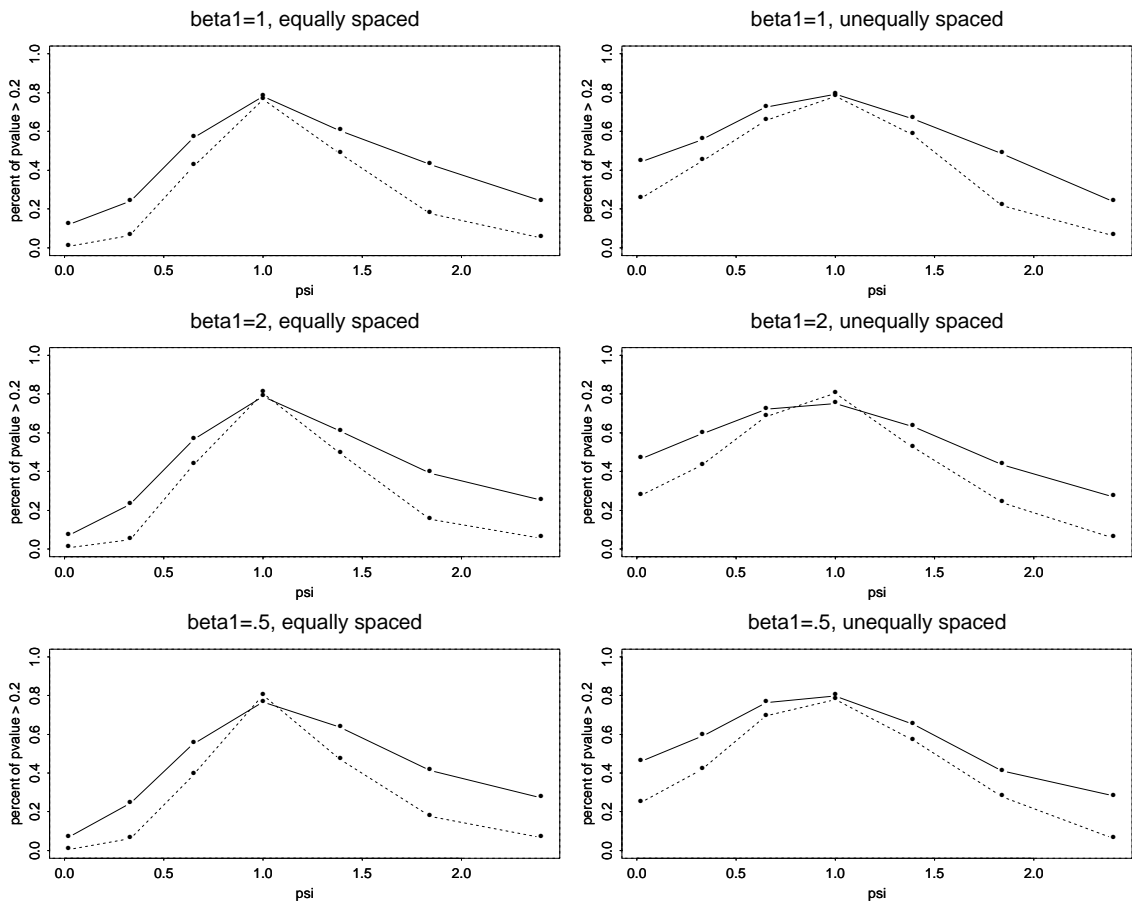
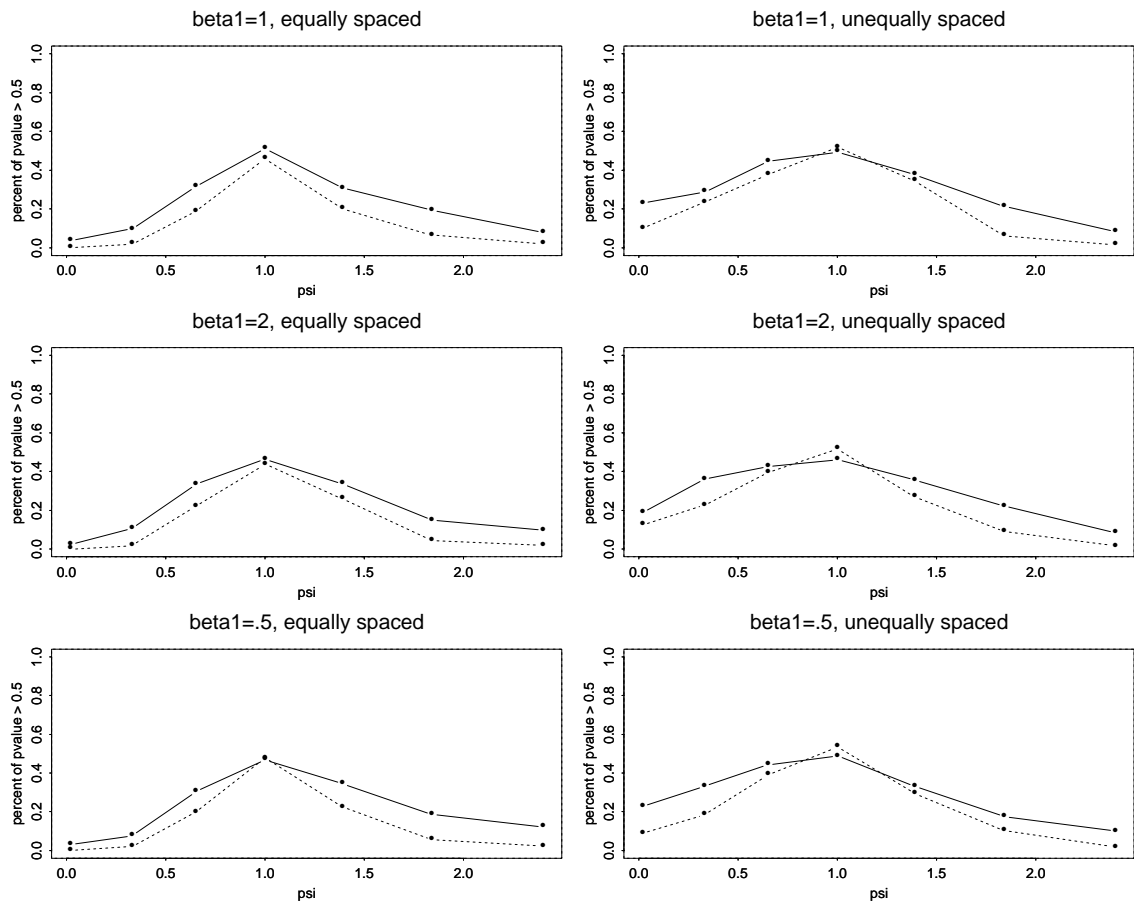*Figure 4.4:* Percentage of likelihood ratio tests of $H : \psi = 1$ versus $K : \psi \neq 1$ which result in a p-value $> .2$

Figure 4.5: Percentage of likelihood ratio tests of $H : \psi = 1$ versus $K : \psi \neq 1$ which result in a p-value $> .5$
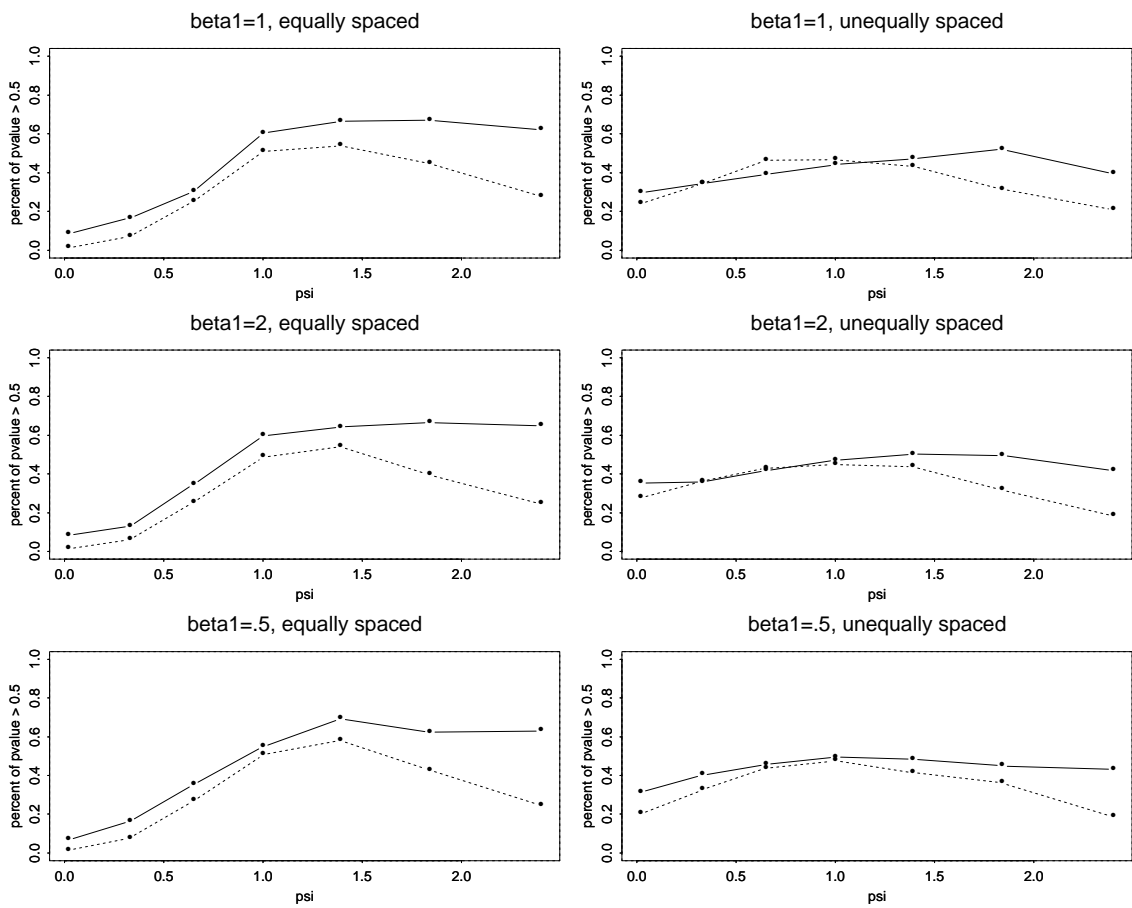
Figure 4.6: Percentage of residual deviance tests assuming a logistic model which result in a *p*-value > .5
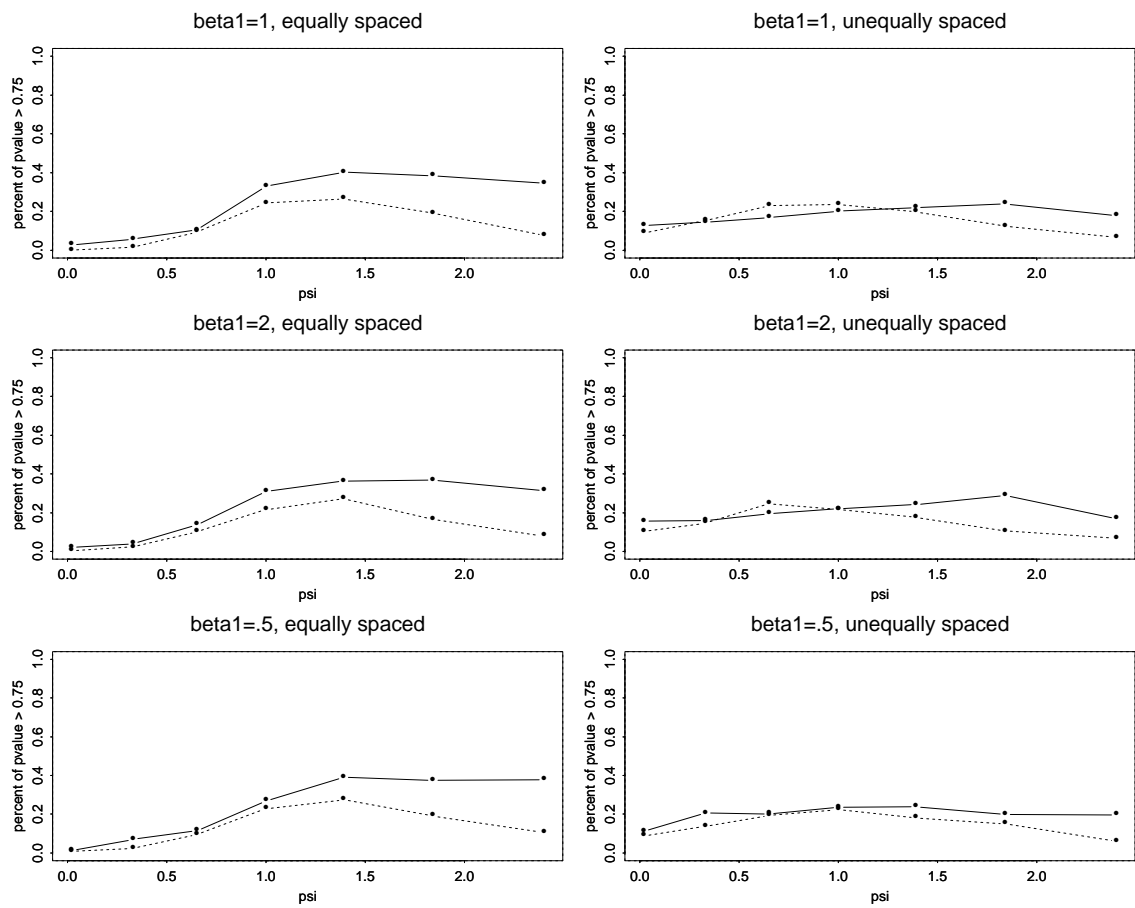
Figure 4.7: Percentage of residual deviance tests assuming a logistic model which result in a
p-value > .75
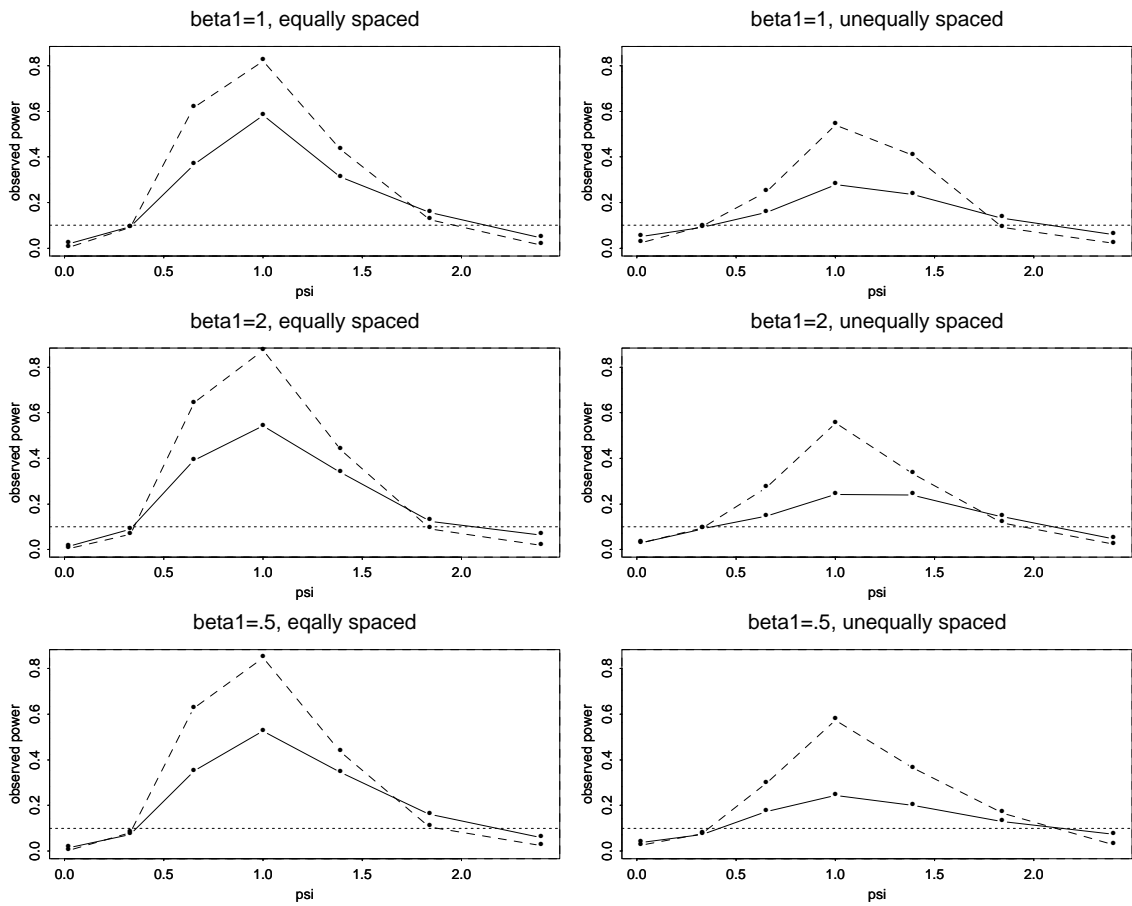
Figure 4.8: Observed power of the link validation test for $H : \Delta_p(\psi) > .1$ versus $K : \Delta_p(\psi) \leq .1$ at significance level $\alpha = .1$
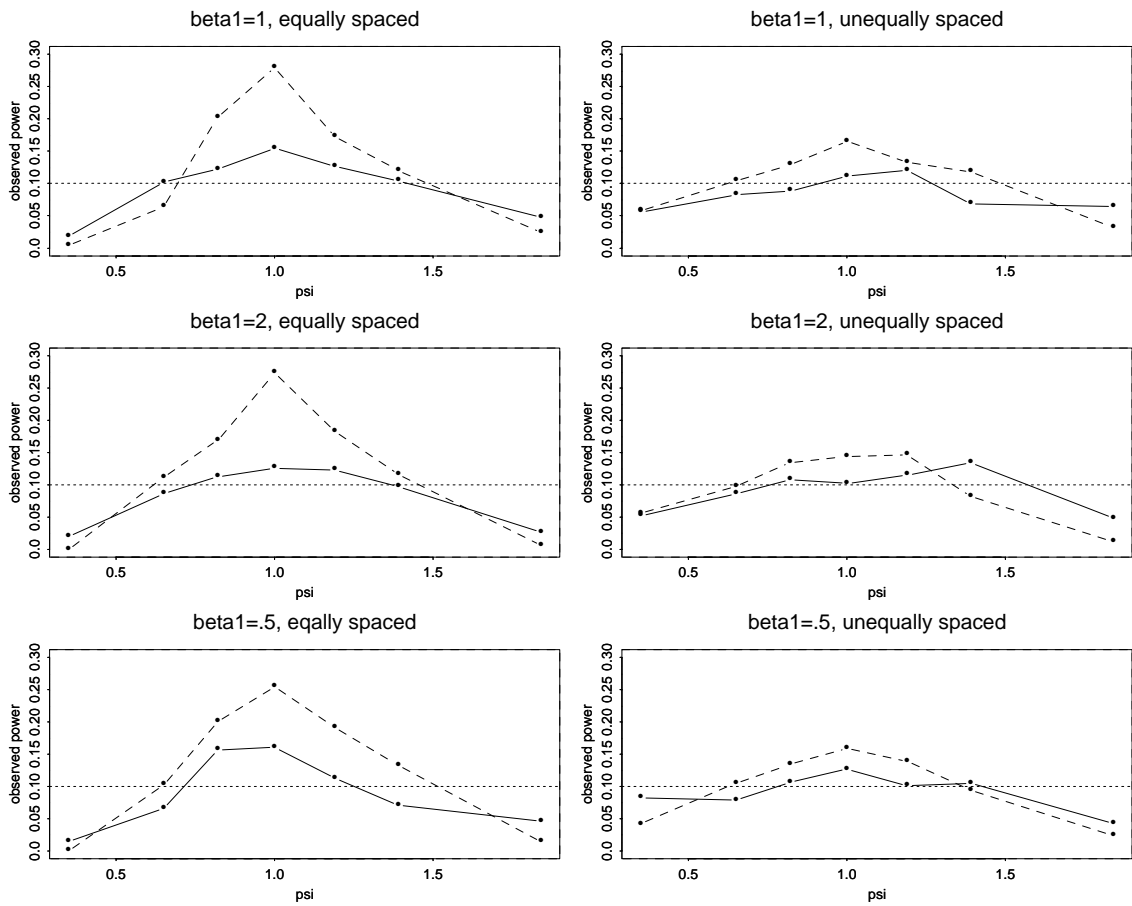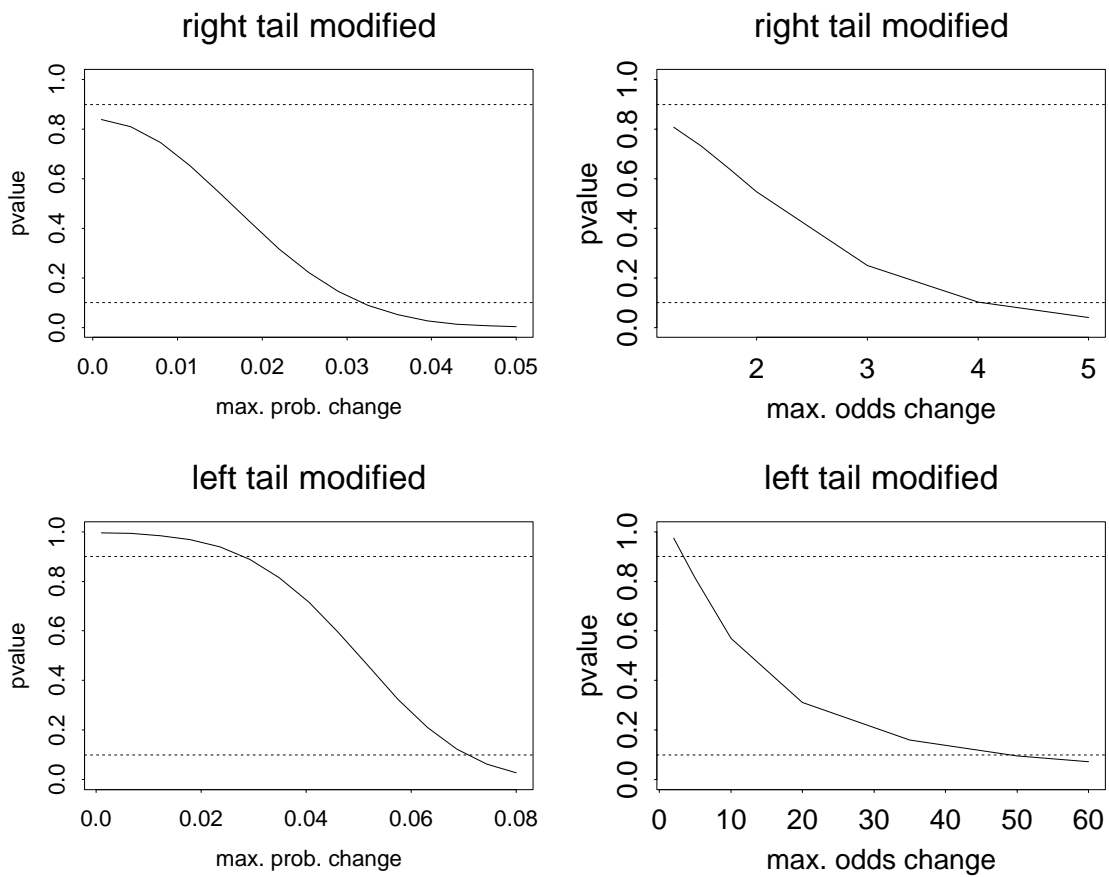
Figure 4.9: Observed power of the link validation test for $H : \Delta_p(\psi) > .05$ versus $K : \Delta_p(\psi) \leq .05$ at significance level $\alpha = .1$

Figure 5.1: P-value curves for assessing the maximal change to estimated success probabilities and odds for the Age to Menarche Data
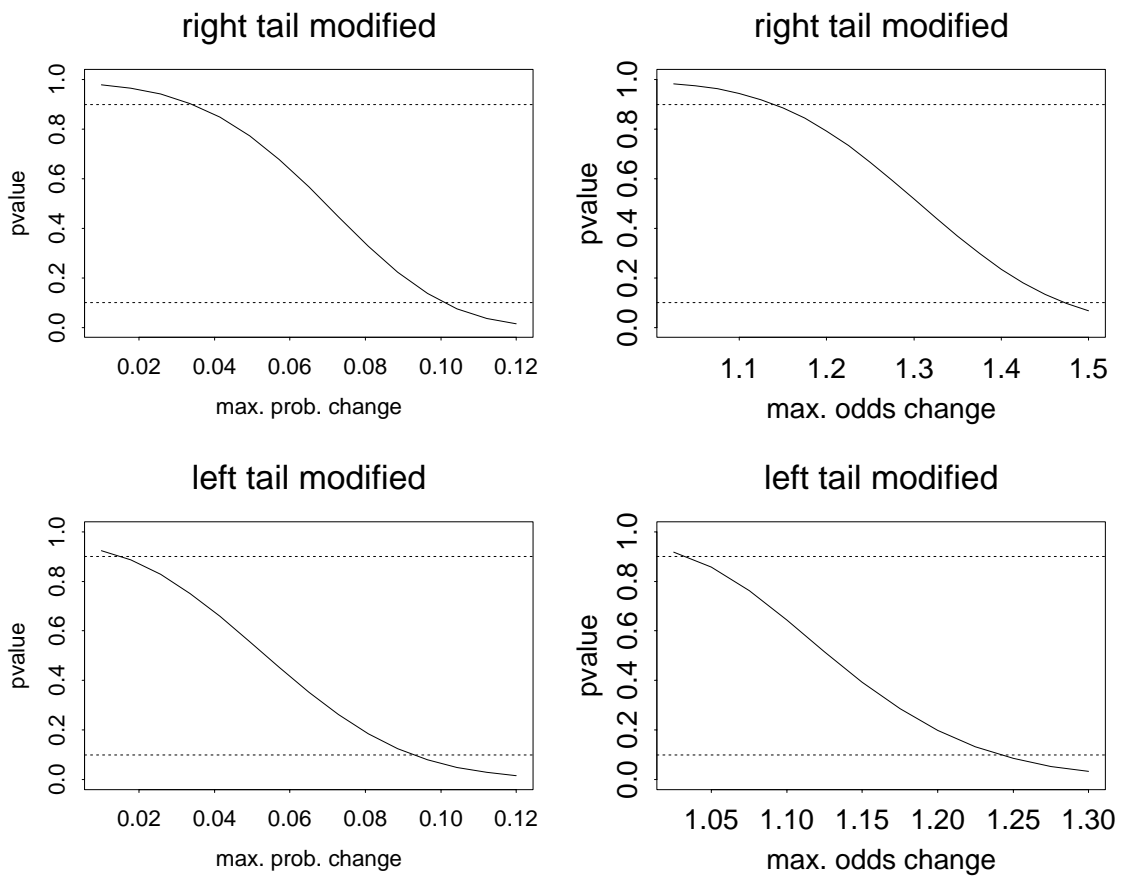
Figure 5.2: *P*-value curves for assessing the maximal change to estimated sucesss probabilities and odds for the Bottle Deposit Data
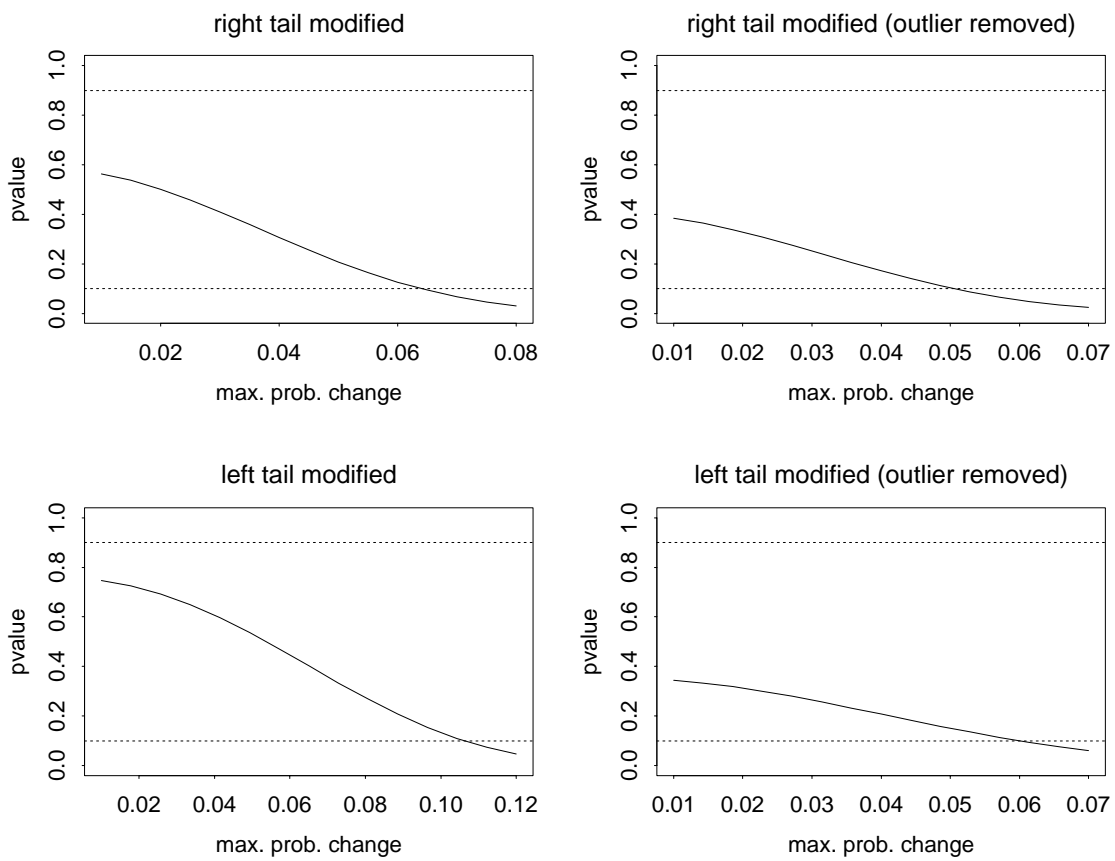
Figure 5.3: *P*-value curves for assessing the maximal change to estimated success probabilities for the complete and outlier removed Flour Beetle Data
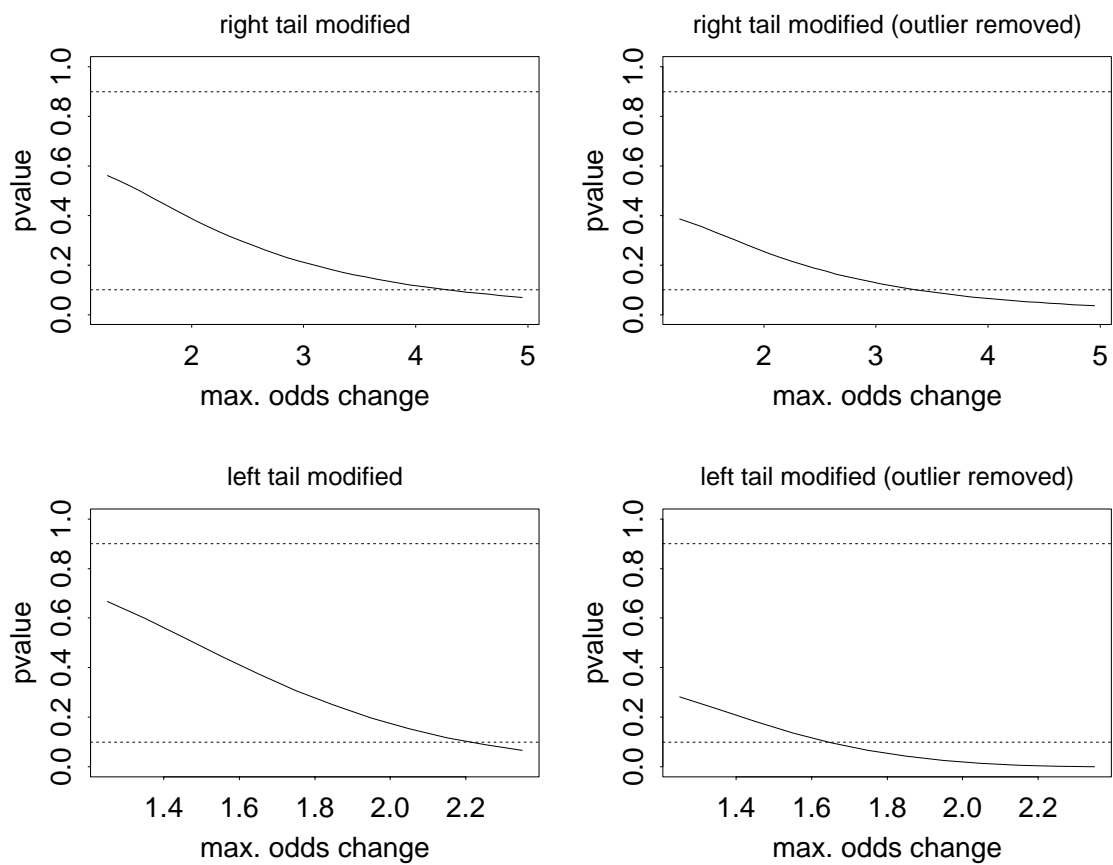
Figure 5.4: *P*-value curves for assessing the maximal change to estimated odds for the complete and outlier removed Flour Beetle Data