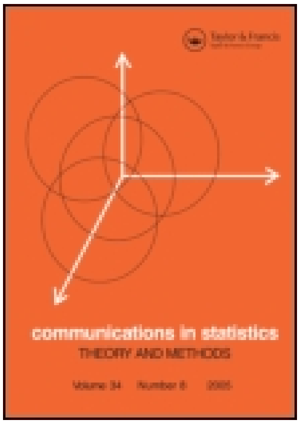


Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

# Norm restricted maximum likelihood estimators for binary regression models with parametric link

Claudia Czado [a]

[a] Department of Mathematics and Statistics, York University, North York, M3J 1P3, Canada

Version of record first published: 27 Jun 2007.

COMMUN. STATIST.—THEORY METH., 22(8), 2259-2274 (1993)

# NORM RESTRICTED MAXIMUM LIKELIHOOD ESTIMATORS FOR BINARY REGRESSION MODELS WITH PARAMETRIC LINK

Claudia Czado
Department of Mathematics and Statistics
York University, North York, Canada, M3J 1P3

*Keywords and phrases: Binary response models, link transformations, logistic regression, EM algorithm, empirical Bayes methods, posterior modes, restricted maximum likelihood*

**Abstract**

Parametric link transformation families have shown to be useful in the analysis of binary regression data since they avoid the problem of link misspecifaction. Inference for these models are commonly based on likelihood methods. Duffy and Santner (1988, 1989) however showed that ordinary logistic maximum likelihood estimators (MLE) have poor mean square error (MSE) behavior in small samples compared to alternative norm restricted estimators. This paper extends these alternative norm restricted estimators to binary regression models with any specified parametric link family. These extended norm restricted MLE's are strongly consistent and efficient under regularity conditions. Finally a simulation study shows that an empiric version of norm restricted MLE's exhibit superior MSE behavior in small samples compared to MLE's with fixed known link.

## 1 INTRODUCTION

Common binary regression models such as logistic or probit regression have been extended to allow for an estimated link by using parametric link transformation families. This avoids the problem of misspecification (see for example Czado and Santner (1992a) and has yielded substantially improved fits in some data sets. Inference for these binary regression models with parametric link is commonly based on likelihood methods. In particular, regression and link parameters are jointly estimated by maximum likelihood. Logistic regression estimates however have shown to have higher MSE in small samples compared to some alternative estimators developed by Duffy and Santner (1988,1989). This paper shows that these alternative estimators can be extended to cover binary regression models with parametric link while maintaining their better MSE behavior.

In binary regression, the observed data is $\{(Y_i, X_i), 1 \le i \le n\}$ with $Y_i$ a 0/1 response and $X_i = (X_{i1}, ..., X_{ip})^t$ a vector of p (possibly stochastic) explanatory variables. Binary regression models with parametric link transformations have the form

$$p(x) = P(Y = 1|x) = F(x^t\beta, \psi) \tag{1.1}$$

where $\{F(\cdot|\psi) : \psi \in \Psi\}$ is a family of cumulative "link" distribution functions and $\beta \in \mathbb{R}^p$ is an unknown regression parameter. If $\{F(\cdot|\psi) : \psi \in \Psi\}$ contains only the logistic distribution, then (1.1) reduces to logistic regression. In general, $\psi$ is called the link parameter and is jointly estimated with the regression parameter $\beta$. In this paper attention is restricted to link distributions $F(z|\psi)$ which are continuously differentiable with respect to z and $\psi$.

Many link distribution families $\{F(\cdot|\psi) : \psi \in \Psi\}$ have been proposed in the literature (see for example Prentice (1975, 1976), Copenhaver and Mielke (1977), Pregibon (1980), Guerrero and Johnson (1982), Aranda-Ordaz (1981), Stukel (1988), Czado and Santner (1992b) and Czado (1992a, 1992b)). Under the assumption that the true link is a member of the specified link family given in (1.1) and regularity conditions, Czado (1989) has shown that $\hat\theta_n = (\hat\psi_n, \hat\beta_n)$, the joint maximum likelihood estimator (MLE) of $\theta = (\psi, \beta)$, is strongly consistent and efficient for fixed and stochastic covariates.

While large sample properties of $\hat\theta_n$ are therefore known, much less is known about its small sample behavior. This leads to the consideration of alternative estimates of $\theta$. This paper introduces a norm restricted MLE of $\theta$ and its empirical version. They can be formally derived by using modifications of Bayes and empirical Bayes methods and are extensions of Duffy and Santner's (1988, 1989) alternative logistic estimators to the case of binary regression with parametric link.

For the large sample theory of the norm restricted MLE's and for the derivation of the empirical version of the norm restricted MLE, it will be necessary to consider stochastic covariates X yielding an extended version of Model (1.1) as follows:

$$\begin{aligned} p(x) = P(Y = 1|X = x) &= F(x^t\beta, \psi) \\ X &\text{ has density } h(x) \end{aligned} \tag{1.2}$$

Thus $\{(Y_i, X_i), 1 \le i \le n\}$ is a multivariate i.i.d sample. Random covariates occur in cohort sampling where the response and explanatory variables are jointly observed but not in experimental settings where covariates are fixed in advance.

The remainder of the paper is orgainized as follows. Section 2 defines norm restricted MLE's for $\theta$ for any specified link family and studies their large sample properties. Section 3 considers an empirical version of the norm restricted MLE's introduced in Section 2. The paper concludes with the results of a simulation study aimed at investigating small sample properties of the alternatives considered. The simulation uses the following parametric link family

$$F(z|\psi_+ = (c_+, k_+)) = 1 - (1 + c_+ \exp(z))^{-k_+} \text{ for } c_+ > 0, k_+ > 0 \tag{1.3}$$

This link family was motivated by Burr (1942) and Prentice(1975, 1976), and used for binary regression with parametric link by Czado and Santner (1992b). $F(z|\psi_+ = (c_+, k_+))$ is positively (negatively) skewed if $k_+ < 1 (> 1)$. The family includes the logistic ($k_+ = 1$) and the extreme minimum value distribution ($k_+ \to \infty$). Moments and other properties of family (1.3) as well as other generalizations of the logistic distribution are given in Balakrishnan and Leung (1988) and Balakrishnan (1992, Chapter 9).

# 2 NORM RESTRICTED MLE'S IN BINARY REGRESSION MODELS WITH PARAMETRIC LINK

To extend the Duffy and Santner's alternative logistic regression estimators to binary regression models with parametric link, let $\mathbf{p}(\theta) = (p_1(\theta), \ldots, p_n(\theta))^t$ with $p_i(\theta) = F(x_i^t\beta, \psi)$ the vector of success probabilities, $l(y, \theta)$ the likelihood and $f(y, \theta)$ the joint density for Model (1.1), i.e.

$$l(y,\theta) = f(y,\theta) = \prod_{i=1}^{n} p_i(\theta)^{y_i}(1 - p_i(\theta))^{1-y_i} \tag{2.1}$$

where $y = (y_1, \ldots, y_n)^t \in [0,1]^n$ and $\theta = (\psi, \beta)$. The mean square error (MSE) of $\hat{p}^M(y)$, the MLE of $\mathbf{p}(\theta)$, is therefore given by

$$MSE_\theta(\hat{p}^M(y)) = \sum_{y \in [0,1]^n} \|\hat{p}^M(y) - p(\theta)\|^2 f(y,\theta). \tag{2.2}$$

It can be argued similarly as in Duffy and Santner (1988) that for the Model (1.1) and for every $\psi \in \Psi$ the MSE of $\hat{p}^M(y)$ is small when the norm of $\beta$ is large, in particular the following holds $\forall$ fixed $\psi$

$$MSE_{\theta_k}(\hat{p}^M(y)) = \sum_{y \in [0,1]^n} \|\hat{p}^M(y) - p(\theta_k)\|^2 f(y,\theta_k) \to 0 \text{ as } k \to \infty, \tag{2.3}$$

where $\theta_k = (\psi, k\beta)$. It is straight forward to see that (2.3) also holds in the case of stochastic covariates (Model (1.3)).

For logistic regression, Duffy and Santner (1988) were able to show that $MSE_\beta(\hat{p}^M(y))$ has a stationary point at $\beta$=0. Their argument relies heavily on the linearity in $\mathbf{p}$ of the score equations and the symmetry of the logistic distribution. Since these properties do not carry over for the general case considered in (1.1) or (1.3), the above result cannot be established in general. However the simulation results presented later indicate that $MSE_\beta(\hat{p}^M(y))$ is larger for $\beta$ small in norm for the cases considered.

Further, experience in analyzing real and simulated data shows that flat likelihood surfaces can occur indicating near nonidentifiability of the parameter $\theta$. In these situations, components of the joint MLE $\hat{\theta}_n$ are driven to $\pm\infty$ thus making model interpretation difficult. This also makes the introduction of norm restricted MLE's of $\theta$ desirable.

General restricted MLE's for $\theta$ can now be defined using a Baysian setup. Assume that $\theta$ has a prior density given by $g(\theta) = g_\beta(\beta) \cdot g_{\psi_1}(\psi_1) \cdots g_{\psi_q}(\psi_q)$. This assumes the prior independence of $\beta$ with each of the link components $\psi_j, j = 1, \ldots, q$. It is further assumed that $g_\beta(\cdot)$ and $g_{\psi_j}(\cdot), j = 1, \ldots, q$ are continuous differentiable densities . Therefore the log posterior likelihood of the data based on Model (1.1) or (1.3), denoted by $l_p(y, \theta)$, is given apart from constants independent of $\theta$ by

$$\begin{aligned} l_p(y,\theta) &= \ln l(y,\theta) + \ln g(\theta) \\ &= \ln l(y,\theta) + \ln g_\beta(\beta) + \sum_{j=1}^{q} \ln g_{\psi_j}(\psi_j) \end{aligned} \tag{2.4}$$

Since $\theta \in \mathbf{R}^{p+q}$, it is computationally difficult to calculate the posterior expected loss, even with respect to square error loss, we have foccused attention on determining the mode of the posterior distribution, denoted by $\hat{\theta}^R$, which is defined by

$$l_p(y, \hat{\theta}^R) = \sup_{\theta \in \mathbf{R}^{p+q}} l_p(y,\theta) \tag{2.5}$$

Assuming stochastic covariates (Model (1.3)) and the same regularity conditions needed for strong consistency and efficiency of $\hat{\theta}_n$, the MLE of $\theta$, plus the log concavity and smoothness of $g(\theta)$, Czado (1989, Theorems 5.2.1-5.2.3) showed that $\hat{\theta}^R$ defined in (2.7) exists asymptotically, is strongly consistent and efficient. As in Duffy and Santner, the estimator $\hat{\theta}^R$ can equivalently by motivated as a restricted MLE by the following proposition

**Proposition 2.1** *Assuming that $g(\cdot)$ and $l(y,\theta)$ are log concave and differentiable at $\hat{\theta}^R$, the estimator $\hat{\theta}^R = (\hat{\psi}_1^R, \ldots, \hat{\psi}_q^R, \hat{\beta}^R)$ solves the following restricted likelihood problem:*

$R$ : *maximize* $\ln l(y,\theta)$ *subject to*

$$\begin{aligned} \ln g_\beta(\beta) &\geq K_\beta \\ \ln g_{\psi_j}(\psi_j) &\geq K_{\psi_j}\ j = 1, \ldots, q \end{aligned} \tag{2.6}$$

*where*

$$\begin{aligned} K_\beta &= \ln g_\beta(\hat{\beta}^R) \\ K_{\psi_j} &= \ln g_{\psi_j}(\hat{\psi}_j^R)\ j = 1, \ldots, q \end{aligned}$$

This proposition is similarly proved as Theorem 3.1 in Duffy and Santner(1988) using a multivariate extension. The smoothness conditions in the proposition insure the existence of the norm restricted estimator $\hat{\theta}^R$. Since densities defined on $\mathbf{R}^p$ converge to zero at the boundaries, (2.6) restricts $\theta$ from becoming too large. Proposition 2.1 allows for the construction of a wide class of restricted MLE's. It is now shown that with the appropiate choice of $g(\theta)$, norm restricted MLE's can be constructed as well. One possible class of prior densities $g(\cdot|\sigma^2)$ for $\theta$ can be given by

$$g(\theta|\sigma^2) = g_\beta(\beta|\sigma_\beta^2) \cdot \prod_{j=1}^{q} g_j(\psi_j|\sigma_j^2) \tag{2.7}$$

where

$$g_\beta(\beta|\sigma_\beta^2) = \frac{1}{(2\pi\sigma_\beta^2)^{p/2}} \exp[-\frac{1}{2\sigma_\beta^2}||\beta||^2] \tag{2.8}$$

$$g_j(\psi_j|\sigma_j^2) = \frac{1}{(2\pi\sigma_j^2)^{1/2}} \exp[-\frac{1}{2\sigma_j^2}(\psi_j - \psi_j^o)^2], \tag{2.9}$$

$\sigma^2 = (\sigma_1^2, \ldots, \sigma_q^2, \sigma_\beta^2)$ are assumed to be known positive constants and $\psi^o = (\psi_1^o, \ldots, \psi_q^o)$ is a specified vector in $\mathbf{R}^q$. This class of priors assumes the apriori independence of the link and regression parameters. The parameter $\beta$ is assumed to be p-variate normal with zero mean vector and covariance matrix $\sigma_\beta^2 I_p$, where $I_p$ denotes the identity matrix of size p, while the prior distributions for $\psi_j, j = 1, \ldots, q$ are independently normal with mean $\psi_j^o$ and variance $\sigma_j^2$ . A reasonable choice for $\psi^o$ would be the value of the link parameter $\psi$ which corresponds to logistic regression or some other commonly used fixed link binary regression model.

Therefore $g(\cdot|\sigma^2)$ is a prior distribution which pulls the regression parameter $\beta$ towards zero and the link parameter $\psi$ towards $\psi^o$. The amount of shrinkage towards these values depends on $\sigma^2$. It is easy to see that this class of prior densities gives norm restricted MLE's, since

$$\begin{aligned} \ln g_\beta(\beta|\sigma_\beta^2) \geq K_\beta &\Leftrightarrow ||\beta||^2 \leq ||\hat{\beta}^R||^2 \\ \ln g_j(\psi_j|\sigma_j^2) \geq K_{\psi_j} &\Leftrightarrow (\psi_j - \psi_j^o)^2 \leq (\hat{\psi}_j^R - \psi_j^o)^2 \end{aligned}$$

where $\hat{\theta}^R = (\hat{\psi}_1^R, \ldots, \hat{\psi}_q^R, \hat{\beta}^R)$ maximizes the log posterior likelihood (2.6) which in this case is given, apart from constants independent of $\theta$, by

$$l_p(y,\theta) = l_p(y,\theta|\sigma^2) =$$
$$l(y,\theta) - \frac{||\beta||^2}{2\sigma_\beta^2} - \sum_{j=1}^{q} \frac{(\psi_j - \psi_j^\circ)^2}{2\sigma_j^2} - \frac{p \cdot \ln(\sigma_\beta^2)}{2} - \sum_{j=1}^{q} \frac{\ln(\sigma_j^2)}{2} \tag{2.10}$$

The choice of $\sigma^2$ is crucial since it govern the degree of shrinkage. This holds since $\hat{\beta}^R$ (or $\hat{\psi}_j^R$) $\to 0$ as $\sigma_\beta^2$ (or $\sigma_j^2$) $\to 0$ and $||\hat{\beta}^R||^2$ (or $(\hat{\psi}_j^R - \psi_j^\circ)^2$) increases as $\sigma_\beta^2$ (or $\sigma_j^2$) increases while all other prior variances remain fixed. This results can be shown similarly as in Duffy and Santner(1988) . However, in general there is little information available to guide the selection of $\sigma^2$. Section 3 proposes an empiric restricted MLE with a data dependent choice of $\sigma^2$.

# 3 EMPIRIC RESTRICTED MLE'S FOR BINARY REGRESSION MODELS WITH PARAMETRIC LINK

To derive the empiric restricted MLE for binary regression with parametric link an i.i.d. sample is needed and therefore Model (1.3) is assumed to hold from now on.

The empirical Bayes estimate of $\sigma^2 = (\sigma_1^2, \ldots, \sigma_q^2, \sigma_\beta^2)$ is given as the maximizer of the marginal likelihood

$$m(y,x|\sigma^2) = \int_{\mathbb{R}^{p+q}} f(y,x|\theta) g(\theta|\sigma^2) d\theta$$

where $f(y,x|\theta)$ is the joint density of $(Y,X) = (Y_1, \ldots, Y_n, X_1, \ldots, X_n)$, i.e.

$$f(y,x|\theta) = \prod_{i=1}^{n} F(x_i^t\beta, \psi)^{y_i}(1 - F(x_i^t\beta,\psi))^{1-y_i} h(x_i) \tag{3.1}$$

and $g(\theta|\sigma^2)$ the prior density given by (2.7). Once the empirical estimate of $\sigma^2$ is determined, the norm restricted MLE $\hat{\theta}^R$ is calculated based on the estimated $\sigma^2$. However, maximization of $m(y,x|\sigma^2)$ over $\sigma^2 \in (0,\infty)$ is computationally infeasible since it would involve a (p+q) dimensional numerical integration.

As one possible alternative the EM algorithm of Dempster, Laird and Rubin (1977) can be applied. Consider the incomplete data to be $(Y,X)$ and the complete data to be $(Y,X,\theta)$ with joint log density

$$d(y,x,\theta|\sigma^2) = \ln f(y,x|\theta) + \ln g(\theta|\sigma^2) \tag{3.2}$$

where $f(y,x|\theta)$ is given by (3.1) and $g(\theta|\sigma^2)$ by (2.7). Ignoring constants independent of $\sigma^2$, (3.2) can be written as

$$d(y,x,\theta|\sigma^2) = -\frac{||\beta||^2}{2\sigma_\beta^2} - \sum_{j=1}^{q} \frac{(\psi_j - \psi_j^\circ)^2}{2\sigma_j^2} - \frac{1}{2}\sum_{j=1}^{q} \ln \sigma_j^2 - \frac{p}{2}\ln(\sigma_\beta^2)$$

it follows that $t(y,x,\theta) = ((\psi_1 - \psi_1^\circ)^2, \ldots, (\psi_q - \psi_q^\circ)^2, ||\beta||^2)$ is a sufficient statistics for $\sigma^2$. Therefore the EM algorithm applied to (3.2) has $s^{th}$ iteration with current guess $\sigma_s^2$ as:

$$\boldsymbol{E-Step}: \quad \text{Estimate } t(y,x,\theta) = ((\psi_1 - \psi_1^\circ)^2, \ldots, (\psi_q - \psi_q^\circ)^2, ||\beta||^2) \text{ by}$$
$$t^s = (t_1^s, \ldots, t_q^s, t_\beta^s) = E(t(y,x,\theta)|Y,X,\sigma_s^2)$$

$$= \frac{1}{m(y,x,\theta|\sigma_s^2)} \int_{\mathbb{R}^{p+q}} t(y,x,\theta) \cdot \exp(d(y,x,\theta|\sigma_s^2))d\theta \qquad (3.3)$$

where integration is understood componentwise.

*M − Step* : Choose $\sigma_{s+1}^2$ to maximize $d(y,x,\theta|\sigma^2)$ with $t(y,x,\theta) = t^s$ over $\sigma^2$.

Since the E-Step (Equation (3.3)) requires a (p+q) dimensional integration, which is computationally impractical, the following multivariate normal approximation of the distribution of $\theta$ given $(Y, X, \sigma_s^2)$ is made. It is multivariate normal with mean $\hat{\theta}_s^R = \hat{\theta}^R(\sigma_s^2)$ and covariance matrix

$$\Sigma^s = \left[-\nabla^2 d(y,x,\theta|\sigma_s^2)\right]^{-1}$$

With this choice the approximating distribution and the actual distribution have identical curvature at $\hat{\theta}_s^R$. Leonard (1972, 1975), Laird (1978) and Duffy and Santner(1989) also use normal approximations in the E-Step. The approach of Laird (1978) was followed here, while Duffy and Santner (1989) use a different estimate for the covariance matrix.

Using this approximation the E-Step becomes

$$t^s = (t_1^s, \ldots, t_q^s, t_\beta^s) \text{ where}$$

$$\left.\begin{array}{rl} t_j^s = & (\hat{\psi}_j^R - \psi_j^o)^2 + \Sigma_j^s \quad j = 1, \ldots, q \\ t_\beta^s = & \|\hat{\beta}_s^R\|^2 + tr(\Sigma_\beta^s) \end{array}\right\} \qquad (3.4)$$

where $(\Sigma_1^s, \ldots, \Sigma_q^s, \Sigma_\beta^s)$ is the diagonal of $\Sigma^s$ and tr(A) denotes the trace of the matrix A.

The M-Step can easily be computed, since

$$\max_{\sigma^2, t(y,x,\theta)=t^s} d(y,x,\theta|\sigma^2) =$$

$$\max_{\sigma^2} \left[ -\frac{t_\beta^s}{2\sigma_\beta^2} - \sum_{j=1}^q \frac{t_j^s}{2\sigma_j^2} - \frac{p}{2}\ln(\sigma_\beta^2) - \sum_{j=1}^q \frac{1}{2}\ln(\sigma_j^2) \right] \qquad (3.5)$$

Elementary calculus shows that the solution to (3.5) is given by

$$\tilde{\sigma}^2 = \left(\frac{t_1^s}{2}, \ldots, \frac{t_q^s}{2}, \frac{t_\beta^s}{p}\right). \qquad (3.6)$$

Using this approximation the $s^{th}$ EM iteration becomes

$$\left.\begin{array}{ll} E - Step: & \text{Estimate } t(y,x,\theta) \text{ by } t^s \text{ defined by (3.4)} \\ M - Step: & \text{Set } \sigma_{s+1}^2 = \tilde{\sigma}^2 \text{ where } \tilde{\sigma}^2 \text{ is given in (3.6)} \end{array}\right\} \qquad (3.7)$$

# 4 SIMULATION RESULTS

## 4.1 INTRODUCTION AND STUDY DESIGN

The main objective of the simulation was to asses the small sample behavior of the empiric restricted MLE's for binary regression models with parametric link using the approximate EM algorithm developed in the previous section. To facilitate this, a particular link family had to be chosen. The link family (1.3) was chosen as example for a simple fexible family. Since for this link family $\psi_+ = (c_+, k_+) > 0$, it is numerically more convenient to specify prior densities for $\psi = (c, k) = (\ln c_+, \ln k_+)$ instead. Further, a somewhat different class of prior distributions than given by (2.8) and (2.9) has been selected and is given as class of prior densities $g(\cdot|\sigma^2)$ for $\theta = (\psi = (c,k), \beta)$ with $c \in \mathbf{R}$ and $k \in \mathbf{R}$:

$$g(\theta|\sigma^2) = g_\beta(\beta|\sigma_\beta^2) \cdot g_c(c|\sigma_c^2) \cdot g_k(k|\sigma_k^2) \tag{4.1}$$

where

$$g_\beta(\beta|\sigma_\beta^2) = \frac{1}{(2\pi\sigma_\beta^2)^{p/2}} \exp[-\frac{1}{2\sigma_\beta^2}||\beta||^2] \tag{4.2}$$

$$g_c(c|\sigma_c^2) = \frac{1}{(2\pi\sigma_c^2)^{1/2}} \exp[-\frac{1}{2\sigma_c^2}(c + \frac{\sigma_c^2}{2})^2] \tag{4.3}$$

$$g_k(k|\sigma_k^2) = \frac{1}{(2\pi\sigma_k^2)^{1/2}} \exp[-\frac{1}{2\sigma_k^2}(k + \frac{\sigma_k^2}{2})^2] \tag{4.4}$$

and $\sigma^2 = (\sigma_c^2, \sigma_k^2, \sigma_\beta^2)$ are assumed to be known positive constants. The prior distributions for c and k are chosen in such a way that $c_+ = \exp(c)$ and $k_+ = \exp(k)$ have lognormal distributions with mean 1, the link value corresponding to logistic regression, and variance $\exp(\sigma_c^2) - 1$ and $\exp(\sigma_k^2) - 1$, respectively.

To derive an empiric restricted MLE for this class of prior distributions, the steps of Section 3 are followed. It is easy to see that for (4.1)-(4.4), the complete data $(Y, X, \theta)$ has joint log density given by

$$d(y,x,\theta|\sigma^2) = -\frac{||\beta||^2}{2\sigma_\beta^2} - \frac{c^2}{2\sigma_c^2} - \frac{k^2}{2\sigma_k^2} - \frac{\sigma_c^2}{8} - \frac{\sigma_k^2}{8} - \frac{p}{2}\ln(\sigma_\beta^2) - \frac{1}{2}\ln(\sigma_c^2) - \frac{1}{2}\ln(\sigma_k^2)$$

Therefore $t(y,x,\theta) = (c^2, k^2, ||\beta||^2)$ is a sufficient statistics for $\sigma^2$. So the approximate EM algorithm takes on the form

$$\left.\begin{array}{ll} \boldsymbol{E - Step:} & \text{Estimate } t(y,x,\theta) = (c^2, k^2, ||\beta||^2) \\ & \text{by } t^s = ((\hat{c}_s^R)^2 + \Sigma_c^s, (\hat{k}_s^R)^2 + \Sigma_k^s, ||\hat{\beta}_s^R||^2 + tr(\Sigma_\beta^s)) \\ \boldsymbol{M - Step:} & \text{Set } \sigma_{s+1}^2 = \tilde{\sigma}^2 \text{where} \\ & \tilde{\sigma}^2 = (2\left[(1+t_c^s)^{1/2} - 1\right], 2\left[(1+t_k^s)^{1/2} - 1\right], \frac{t_\beta^s}{p}) \end{array}\right\} \tag{4.5}$$

The approximate EM algorithm (Equation (4.5)) used below was implemented by initializing $\sigma^2 = (1,1,1)$ and $\theta$ at the true underlying parameter value of the simulation. The following inequality was used as stopping criteria for the iterative procedure:

$$||\sigma_s^2 - \sigma_{s+1}^2||^2 < .05. \tag{4.6}$$

As performance measure for the simulation, estimates of bias and mean square error (MSE) of the parameter and success probability estimates were calculated. The empiric restricted MLE's, denoted by EMLE, calculated using the approximate EM algorithm (4.5) was further compared to two other estimation procedures. One method was maximum likelihood estimation of $(c_+, \beta)$ where $k_+$ is fixed at the true underlying value to assess the cost of having to estimate $k_+$ and the other method was maximum likelihood estimation of $(c_+, \beta)$ when $k_+$ is set to 1; this corresponds the logistic regression with intercept $c = \ln c_+$. The first method will be denoted by MLE($k_+$) while the second one by LRE. LRE was included since it corresponds to the standard method for binary regression data. Joint MLE of $\theta = (c, k, \beta)$ was not included since near nonidentifiability occurred when attempted causing numerical problems in the simulation. This is a more general problem for binary regression models with estimated link since the additional link parameter allows more easily for flat likelihood surfaces, thus nonidentifiabilty, compared to a fixed link model. Simulation results concerning restricted MLE's as defined in Section 2 are also not presented since they require a sensible choice of $\sigma^2$ apriori which in general is impossible.

It should also be noted that comparing performance of regression parameters $(c_+, \beta)$ derived from LRE to the corresponding one from EMLE or MLE($k_+$) is less appropiate than comparing the performance of the success probability estimates. The regression parameters depend heavily on the link function used for estimation whereas models with different links can only be unambiguously compared in their estimation of true probabilities. This is in sharp contrast to the use of transformations in the linear model where there has been much controversy in the determination of a common basis to evaluate the performance of estimation procedures (see Hinkley and Runger (1984) and Bickel and Doksum (1981)). It should be noted that the dependence of the regression parameters on the link parameter remains even if an orthogonalized ( see Czado and Santner (1992b)) or more general a standardized link family (see Czado (1992a, 1992b)) which include Stukel's (1988) link family is used, since only local parameter orthogonality can be achieved in this problem.

For the simulation the following binary regression model with parametric link, random covariates and normal prior $g(\cdot|\sigma^2)$ was chosen

**(i) Binary Response**

$$p(x) = P(Y = 1|X = x) = F(x^t\beta|\psi_+ = (c_+, k_+)) = 1 - (1 + c_+ \exp(x^t\beta))^{-k_+}$$

**(ii) Random Covariate**

$X = (X_1, \ldots, X_p)^t$ i.i.d with $X_i$ uniformly distributed over [-1,1].

**(iii) Prior Distribution**

Prior density for $\theta = (c, k, \beta) = (\ln c_+, \ln k_+, \beta)$ is specified by (4.1)-(4.4).

EMLE will give estimates of $\theta = (c, k, \beta)$, therefore $(c_+, k_+)$ are estimated by $(\exp(\hat{c}), \exp(\hat{k}))$ where $(\hat{c}, \hat{k})$ are estimates of $(c, k)$.

The number p of random covariates was set to 5. Two types of departure from logistic link, kurtosis and skewness, were studied and therefore the following three different choices of the true underlying link parameters $(c_+, k_+)$ were made

| *Case* | $c_+$ | $k_+$ |
|---|---|---|
| *logistic* | 1.0 | 1.0 |
| *heavier tail* | 1.7 | .8 |
| *skewed to the left* | .9 | 1.7 |

Figures 4.1 and 4.2 show the corresponding distribution and density functions, respectively.

For each of the three true link combinations $(c_+, k_+)$, four different choices of $\beta \in \mathbf{R}^5$ were made. For one choice of $\beta$, all components are equal (equal beta case) while for the second choice components were set in an increasing order on a straight line with $\beta_5 = 2.5\beta_1$ (increasing beta case). For each of the two cases two different $\beta$ values were selected to correspond to a central case (95 % of all success probabilities are in [.25,.75]) and a noncentral case (50 % of all success probabilities are in [.25,.75]). The central case is denoted by CEN=95% and the noncentral case by CEN=50%. The particular value of $\beta$ was determined empirically based on a sample of size 500, since no easy closed form distribution of sums of uniform random variables exist. Since the results from the increasing beta case do not differ qualitatively from the equal beta case, results of the former will be omitted. Table 4.1 gives the particular choices of $\beta \in \mathbf{R}^5$ and the observed proportion of success probabilities in [.25,.75] for the equal beta case.

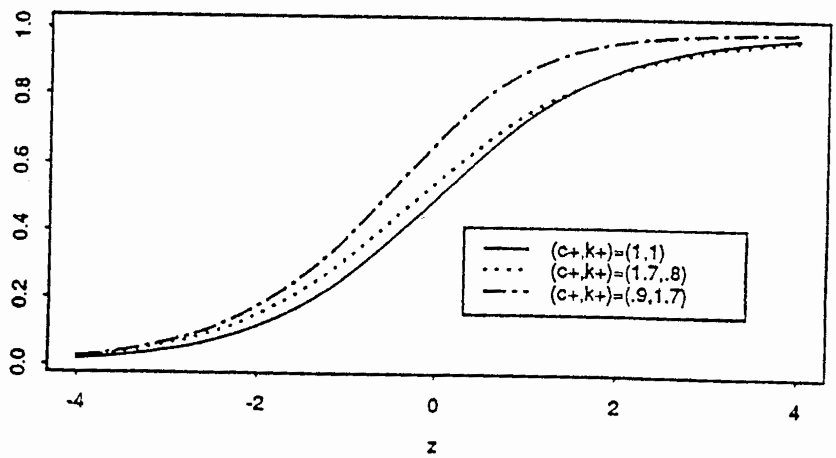


**FIG. 4.1:** True Link Distributions

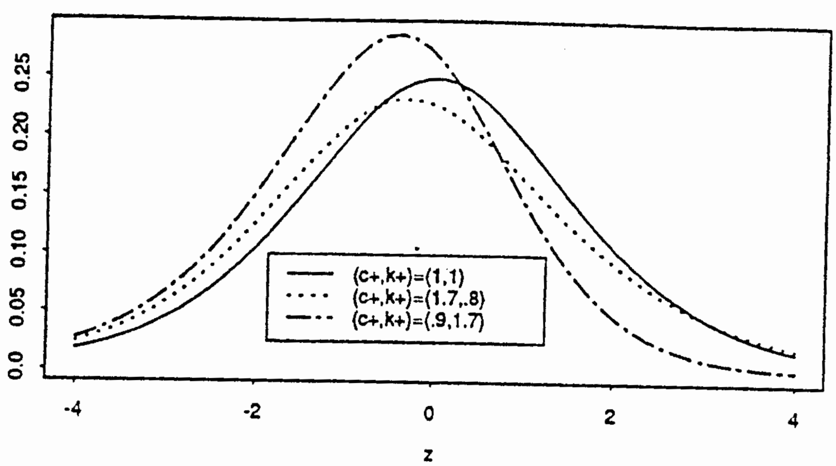


**FIG. 4.2:** True Link Densities

Two sample sizes were investigated, N=50 using 1000 replications and N=200 using 250 replications.

To evaluate performance of the success probability estimates for some specified covariate values, 243 points in design space $[-1,1]^5$ were fixed corresponding to all possible combinations of {-1,0,1}for each of the five components. The mean, median, minimum and

TABLE 4.1: Choice of $\beta \in \mathbf{R}^5$ for the Equal Beta Case

| Case $(c_+, k_+)$ | CEN | Observed CEN | $\beta_i$ |
|---|---|---|---|
| *logistic* (1,1) | .95 | .9443 | 0.450 |
| | .50 | .5202 | 1.180 |
| *heavier tail* (1.7 ,.8) | .95 | .9436 | 0.465 |
| | .50 | .5134 | 1.290 |
| *skewed to the left* (.9,1.7) | .95 | .9475 | 0.160 |
| | .50 | .5294 | 0.880 |

TABLE 4.2: Estimated Mean Prior Variances

| $(c_+, k_+)$ $CEN$ | | N | $\hat{\sigma}^2_\beta$ | $\hat{\sigma}^2_c$ | $\hat{\sigma}^2_k$ |
|---|---|---|---|---|---|
| (1.,1.) | .95 | 50 | .42(.01) | .16(.00) | .18(.00) |
| | | 200 | .24(.01) | .12(.00) | .14(.00) |
| | .50 | 50 | 1.3(.03) | .18(.00) | .20(.00) |
| | | 200 | 1.3(.03) | .16(.01) | .20(.01) |
| (1.7,.8) | .95 | 50 | .39(.01) | .18(.00) | .15(.00) |
| | | 200 | .20(.01) | .14(.00) | .11(.00) |
| | .50 | 50 | 1.3(.03) | .20(.00) | .17(.00) |
| | | 200 | 1.2(.03) | .17(.01) | .16(.01) |
| (.9,1.7) | .95 | 50 | .26(.00) | .14(.00) | .31(.00) |
| | | 200 | .08(.00) | .12(.00) | .26(.00) |
| | .50 | 50 | .86(.01) | .15(.00) | .37(.01) |
| | | 200 | .79(.01) | .15(.01) | .39(.01) |

The prior variances $\sigma^2_\beta, \sigma^2_c$ and $\sigma^2_k$ were set to 1.

maximum values of bias and MSE of the success probabilities were calculated. For the bias the absolute mean value was also recorded.

Computations were done on an IBM 3090 at Cornell University using the NAG FORTRAN routine E04GFC for all maximizations.

## 4.2 PERFORMANCE RESULTS FOR PARAMETER ESTIMATES

Table 4.2 gives the mean prior variance estimates of $\sigma^2 = (\sigma^2_c, \sigma^2_k, \sigma^2_\beta)$ including estimated standard errors in parentheses for EMLE. It can be seen that $E(\hat{\sigma}^2_\beta)$ is large when the true $\beta$ is large. $E(\hat{\sigma}^2_\beta)$ decreases for N=200 since the prior effect becomes smaller.

For the logistic case one can see that $E(\hat{\sigma}^2_c) \approx E(\hat{\sigma}^2_k)$ reflecting $c_+ = k_+ = 1$. Further, the values of $E(\hat{\sigma}^2_c)$ and $E(\hat{\sigma}^2_k)$ are small, which means that the EMLE's shrink towards the

**TABLE 4.3:** Estimated Bias of $(\hat{c}_+, \hat{k}_+,)$ and $\hat{\beta}_1$ for EMLE, MLE($k_+$) and LRE

| $(c_+, k_+)$ *CEN N* | Estimator | $\hat{k}_+$ | $\hat{c}_+$ | $\hat{\beta}_1$ |
|---|---|---|---|---|
| (1,1) .95 50 | $EMLE$ | .06(.00) | −.08(.00) | −.16(.01) |
| | $MLE(1) = LRE$ | — | .06(.01) | .11(.02) |
| .95 200 | $EMLE$ | .06(.00) | −.08(.00) | −.09(.01) |
| | $MLE(1) = LRE$ | — | −.00(.01) | .03(.02) |
| .5 50 | $EMLE$ | .13(.01) | −.12(.00) | −.32(.02) |
| | $MLE(1) = LRE$ | — | .14(.02) | .33(.03) |
| .5 200 | $EMLE$ | .13(.01) | −.14(.01) | −.10(.02) |
| | $MLE(1) = LRE$ | — | .02(.01) | .08(.02) |
| (1.7,.8) .95 50 | $EMLE$ | .36(.01) | −.73(.00) | −.21(.01) |
| | $MLE(.8)$ | — | .26(.03) | .09(.02) |
| | $LRE$ | — | −.36(.02) | .04(.02) |
| .95 200 | $EMLE$ | .35(.01) | −.72(.00) | −.15(.01) |
| | $MLE(.8)$ | — | .04(.02) | .01(.02) |
| | $LRE$ | — | −.47(.01) | −.03(.02) |
| .5 50 | $EMLE$ | .40(.01) | −.78(.00) | −.44(.02) |
| | $MLE(.8)$ | — | .44(.05) | .39(.03) |
| | $LRE$ | — | −.29(.03) | .25(.03) |
| .5 200 | $EMLE$ | .39(.01) | −.78(.01) | −.23(.02) |
| | $MLE(.8)$ | — | .04(.02) | .09(.02) |
| | $LRE$ | — | −.50(.01) | −.02(.02) |
| (.9, 1.7) .95 50 | $EMLE$ | −.20(.01) | .16(.00) | −.07(.01) |
| | $MLE(1.7)$ | — | .13(.01) | .02(.02) |
| | $LRE$ | — | 1.6(.05) | .06(.02) |
| .95 200 | $EMLE$ | −.20(.01) | .16(.00) | −.05(.01) |
| | $MLE(1.7)$ | — | .02(.01) | .00(.01) |
| | $LRE$ | — | 1.1(.02) | .04(.02) |
| .5 50 | $EMLE$ | −.14(.01) | .09(.00) | −.23(.01) |
| | $MLE(1.7)$ | — | .22(.02) | .23(.03)) |
| | $LRE$ | — | 2.0(.08) | .44(.03) |
| .5 200 | $EMLE$ | −.08(.01) | .06(.01) | −.07(.02) |
| | $MLE(1.7)$ | — | .02(.01) | .03(.02) |
| | $LRE$ | — | 1.2(.02) | .21(.02) |

true parameter values $(c_+, k_+) = (1, 1)$. Since the true $\beta$ is not zero, $E(\hat{\sigma}^2_\beta)$ is larger than $E(\hat{\sigma}^2_c)$ and $E(\hat{\sigma}^2_k)$.

For the heavier tail case, one has $E(\hat{\sigma}^2_c) \approx E(\hat{\sigma}^2_k)$, which does not reflect the true $(c_+, k_+)$. It should be noted that since this distribution does not only has heavier tails but is also more concentrated around zero, where this link is not so different than the logistic link, it is difficult for EMLE to identify the correct $(c_+, k_+)$.

For the skewed to the left case, one has $E(\hat{\sigma}^2_c) < E(\hat{\sigma}^2_k)$ reflecting that $c_+ = .9$ is closer to 1 than $k_+ = 1.7$ is to 1. In all cases the approximate EM algorithm (Equation (4.5)) takes about three iterations on the average to converge from its starting point.

Table 4.3 lists the estimated bias of the parameter estimates of $(c_+, k_+)$ and $\beta_1$ for EMLE, MLE($k_+$) and LRE. Results for $\beta_2, \ldots, \beta_5$ have been omitted since they are similar to the one for $\beta_1$. Estimated standard errors are given in parentheses. Further note that in the case $(c_+, k_+) = (1, 1)$ one has MLE(1)=LRE.

**TABLE 4.4:** Estimated MSE of $(\hat{c}_+, \hat{k}_+,)$ and $\hat{\beta}_1$ for EMLE, MLE($k_+$) and LRE

| $(c_+, k_+)$ *CEN* *N* | Estimator | $\hat{k}_+$ | $\hat{c}_+$ | $\hat{\beta}_1$ |
|---|---|---|---|---|
| (1,1) .95 50 | *EMLE* | .02(.00) | .02(.00) | .15(.01) |
| | *MLE*(1) = *LRE* | — | .19(.02) | .46(.03) |
| .95 200 | *EMLE* | .01(.00) | .01(.00) | .05(.00) |
| | *MLE*(1) = *LRE* | — | .02(.00) | .07(.01) |
| .5 50 | *EMLE* | .04(.00) | .03(.00) | .37(.01) |
| | *MLE*(1) = *LRE* | — | .55(.19) | 1.0(.08) |
| .5 200 | *EMLE* | .04(.01) | .03(.00) | .10(.01) |
| | *MLE*(1) = *LRE* | — | .03(.00) | .12(.01) |
| (1.7,.8) .95 50 | *EMLE* | .16(.01) | .54(.00) | .16(.01) |
| | *MLE*(.8) | — | .98(.12) | .54(.03)) |
| | *LRE* | — | .46(.03) | .44(.03) |
| .95 200 | *EMLE* | .13(.00) | .52(.00) | .06(.00) |
| | *MLE*(.8) | — | .08(.01) | .08(.01) |
| | *LRE* | — | .26(.01) | .07(.01) |
| .5 50 | *EMLE* | .20(.01) | .62(.01) | .45(.02) |
| | *MLE*(.8) | — | 3.3(.90) | 1.4(.21) |
| | *LRE* | — | 1.0(.19) | 1.1(.17) |
| .5 200 | *EMLE* | .17(.01) | .62(.01) | .14(.01) |
| | *MLE*(.8) | — | .13(.01) | .15(.02) |
| | *LRE* | — | .30(.01) | .12(.01) |
| (.9, 1.7) .95 50 | *EMLE* | .13(.00) | .03(.00) | .09(.01) |
| | *MLE*(1.7) | — | .20(.04) | .34(.03) |
| | *LRE* | — | 4.5(.63) | .50(.04) |
| .95 200 | *EMLE* | .06(.00) | .03(.00) | .02(.00) |
| | *MLE*(1.7) | — | .01(.00) | .05(.00) |
| | *LRE* | — | 1.4(.05) | .07(.01) |
| .5 50 | *EMLE* | .12(.01) | .01(.00) | .25(.01) |
| | *MLE*(1.7) | — | .58(.18) | .64(.05) |
| | *LRE* | — | 10.(2.8) | 1.0(.08) |
| .5 200 | *EMLE* | .05(.01) | .01(.00) | .07(.01) |
| | *MLE*(1.7) | — | .02(.00) | .07(.01) |
| | *LRE* | — | 1.7(.07) | .15(.01) |

In general, the bias is reduced as sample size or as centrality (CEN) is increased. All estimation methods recognize the equality of the $\beta$ components reasonably well. For the alternative estimator EMLE the estimated bias of $\beta$ is negative indicating an overshrinking towards zero. MLE($k_+$) estimates have lower or equal estimated absolute bias values than for EMLE indicating the cost in having to estimate $k_+$. However, MLE($k_+$) overestimates $(c_+, \beta)$ and this overestimation increases as CEN or sample size decreases. This is analog to similar findings for logistic regression (see Duffy and Santner (1989)). In small noncentral cases, the magnitude of the absolute bias of $\beta$ is the same for EMLE and MLE($k_+$). Further note that EMLE overestimates $k_+$ and underestimate $(c_+, \beta)$ in the logistic and heavier tail case, while $(k_+, \beta)$ is underestimated and $c_+$ overestimated in the skewed to the left case,

TABLE 4.5: Estimated Bias of Success Probabilities for EMLE, MLE($k_+$) and LRE

| $(c_+, k_+)$ *CEN* *N* | Estimator | Min | Mean | Median | \|Mean\| | Max |
|---|---|---|---|---|---|---|
| (1,1) .95 50 | $EMLE$ | −.0047 | −.0040 | −.0044 | .0040 | −.0013 |
| | $MLE(1) = LRE$ | −.0334 | −.0023 | −.0032 | .0102 | .0321 |
| .95 200 | $EMLE$ | −.0037 | −.0016 | −.0025 | .0024 | .0037 |
| | $MLE(1) = LRE$ | −.0134 | −.0023 | −.0022 | .0040 | .0083 |
| .5 50 | $EMLE$ | .0000 | .0039 | .0047 | .0039 | .0085 |
| | $MLE(1) = LRE$ | −.0547 | .0007 | .0002 | .0166 | .0539 |
| .5 200 | $EMLE$ | −.0041 | .0020 | −.0001 | .0043 | .0095 |
| | $MLE(1) = LRE$ | −.0204 | .0005 | .0002 | .0060 | .0210 |
| (1.7,.8) .95 50 | $EMLE$ | −.0248 | −.0052 | −.0087 | .0158 | .0284 |
| | $MLE(.8)$ | −.0361 | −.0038 | −.0051 | .0115 | .0372 |
| | $LRE$ | −.0326 | −.0020 | −.0035 | .0114 | .0408 |
| .95 200 | $EMLE$ | −.0236 | −.0009 | −.0043 | .0166 | .0347 |
| | $MLE(.8)$ | −.0145 | −.0010 | −.0010 | .0039 | .0093 |
| | $LRE$ | −.0141 | .0002 | .0002 | .0047 | .0144 |
| .5 50 | $EMLE$ | −.0285 | −.0010 | −.0122 | .0189 | .0272 |
| | $MLE(.8)$ | −.0546 | −.0013 | −.0024 | .0164 | .0502 |
| | $LRE$ | −.0556 | −.0004 | −.0012 | .0170 | .0515 |
| .5 200 | $EMLE$ | −.0310 | −.0008 | −.0132 | .0208 | .0303 |
| | $MLE(.8)$ | −.0236 | −.0011 | −.0002 | .0059 | .0204 |
| | $LRE$ | −.0284 | .0002 | .0026 | .0076 | .0203 |
| (.9, 1.7) .95 50 | $EMLE$ | −.0199 | −.0117 | −.0118 | .0117 | −.0012 |
| | $MLE(1.7)$ | −.0459 | −.0141 | −.0131 | .0149 | .0142 |
| | $LRE$ | −.0502 | −.0173 | −.0161 | .0179 | .0189 |
| .95 200 | $EMLE$ | −.0106 | −.0041 | −.0042 | .0043 | .0031 |
| | $MLE(1.7)$ | −.0185 | −.0034 | −.0028 | .0057 | .0076 |
| | $LRE$ | −.0204 | −.0049 | −.0043 | .0065 | .0079 |
| .5 50 | $EMLE$ | −.0199 | −.0103 | −.0147 | .0104 | .0004 |
| | $MLE(1.7)$ | −.0603 | −.0089 | −.0079 | .0170 | .0294 |
| | $LRE$ | −.0615 | −.0114 | −.0130 | .0171 | .0321 |
| .5 200 | $EMLE$ | −.0072 | −.0034 | −.0049 | .0036 | .0006 |
| | $MLE(1.7)$ | −.0255 | −.0042 | −.0036 | .0062 | .0176 |
| | $LRE$ | −.0270 | −.0072 | −.0093 | .0102 | .0207 |

allowing for compensation at the success probability scale. This is in sharp contrast to the MLE($k_+$) estimates which overestimates all parameters.

Finally, the LRE estimates are more biased in the skewed to the left case than the heavier tail case confirming earlier results of Czado and Santner (1992a) that skewness has larger effects than tail behavior when the link is misspecified.

Turning now to the MSE of the parameter estimates; results are presented in Table 4.4. In general, MSE decreases as either sample size or centrality increases. For small samples, the estimated MSE for the alternative estimates EMLE are roughly 1/3 compared to those derived from MLE($k_+$) and LRE giving strong evidence towards the variance stabilizing effect of the alternative estimation procedure. For the larger sample size this effect is washed out, estimated MSE's are of equal magnitude, and asymptotics takes over.

TABLE 4.6: Estimated MSE of Success Probabilities for EMLE, MLE($k_+$) and LRE

| $(c_+, k_+)$ $CEN$ $N$ | Estimator | Min | Mean | Median | Max |
|---|---|---|---|---|---|
| (1,1) .95 50 | $EMLE$ | .0004 | .0039 | .0044 | .0050 |
| | $MLE(1) = LRE$ | .0085 | .0454 | .0480 | .0691 |
| .95 200 | $EMLE$ | .0001 | .0010 | .0011 | .0013 |
| | $MLE(1) = LRE$ | .0015 | .0110 | .0112 | .0203 |
| .5 50 | $EMLE$ | .0000 | .0024 | .0022 | .0053 |
| | $MLE(1) = LRE$ | .0003 | .0359 | .0414 | .0804 |
| .5 200 | $EMLE$ | .0000 | .0008 | .0006 | .0016 |
| | $MLE(1) = LRE$ | .0000 | .0080 | .0088 | .0249 |
| (1.7,.8) .95 50 | $EMLE$ | .0007 | .0041 | .0044 | .0049 |
| | $MLE(.8)$ | .0083 | .0466 | .0463 | .0729 |
| | $LRE$ | .0082 | .0464 | .0469 | .0719 |
| .95 200 | $EMLE$ | .0003 | .0012 | .0012 | .0017 |
| | $MLE(.8)$ | .0013 | .0112 | .0107 | .0224 |
| | $LRE$ | .0013 | .0112 | .0108 | .0218 |
| .5 50 | $EMLE$ | .0000 | .0027 | .0030 | .0057 |
| | $MLE(.8)$ | .0006 | .0362 | .0373 | .0801 |
| | $LRE$ | .0006 | .0362 | .0386 | .0799 |
| .5 200 | $EMLE$ | .0000 | .0012 | .0016 | .0018 |
| | $MLE(.8)$ | .0000 | .0083 | .0075 | .0242 |
| | $LRE$ | .0000 | .0083 | .0081 | .0242 |
| (.9, 1.7) .95 50 | $EMLE$ | .0038 | .0053 | .0054 | .0055 |
| | $MLE(1.7)$ | .0074 | .0479 | .0491 | .0703 |
| | $LRE$ | .0075 | .0490 | .0501 | .0734 |
| .95 200 | $EMLE$ | .0008 | .0011 | .0011 | .0011 |
| | $MLE(1.7)$ | .0012 | .0117 | .0116 | .0213 |
| | $LRE$ | .0012 | .0118 | .0116 | .0223 |
| .5 50 | $EMLE$ | .0000 | .0032 | .0037 | .0059 |
| | $MLE(1.7)$ | .0004 | .0372 | .0336 | .0831 |
| | $LRE$ | .0005 | .0379 | .0354 | .0874 |
| .5 200 | $EMLE$ | .0000 | .0007 | .0008 | .0014 |
| | $MLE(1.7)$ | .0000 | .0088 | .0073 | .0243 |
| | $LRE$ | .0000 | .0091 | .0072 | .0270 |

## 4.3 PERFORMANCE RESULTS FOR EVENT PROBABILITY ESTIMATES

Table 4.5 gives estimated minimum, mean, median, absolute mean and maximum bias values for the probabilities at 243 x points identified in Section 4.1. The minimum and maximum values based on EMLE are lowest except for the large sample heavier tail cases compared to all the other estimation procedures. They are significantly lower for the small sample sizes ranging from 1/7 to 1/24 in the logistic case compared to MLE(1), from 3/4 to 1/2 in the heavier tail case compared to MLE(.8) and from 1/2 to 1/80 in the skewed to the left case compared to MLE(1.7). For the large sample size the reduction is around 1/2 in the logistic case and in the skewed to the left case. The heavier tail case is special since the EM algorithm had difficulties identifying the true underlying link, resulting in at most doubling the absolute minimum and maximum bias value compared to MLE(1.7).

The corresponding estimated MSE results for the success probability estimates are given in Table 4.6. Knowing that EMLE is significantly less biased on the probability scale and that the alternative estimation procedures are more stable with regard to variances, one expects a very significant reduction of the estimated maximum MSE values compared to those of the MLE($k_+$) ones, which is confirmed in Table 4.6.

In detail, the reduction for EMLE in all cases considered is about 1/16 compared to MLE($k_+$). The MSE values for LRE are about the same magnitude than the ones for MLE($k_+$).

## 4.4 SUMMARY AND CONCLUSIONS

As in logistic regression, MLE based on a fixed true link possibly different than the logistic link overestimates on the average the regression parameters. This overestimation increases as centrality increases or sample size decreases. Further, MSE are high both on the parameter as well as on the probability scale in small samples. In contrast, the empiric norm restricted estimators exhibit superior MSE behavior compared to the MLE based on a fixed true link. This shows that as in logistic regression norm restricted estimators are valuable alternatives to MLE's in small samples for binary regression models with parametric link.

## ACKNOWLEDGEMENTS

This research was supported, in part, by the National Supercomputing Facility at Cornell University, a resource of the Center for Theory and Simulation in Science and Engineering, which is funded in part by the National Science Foundation, New York State, and the IBM Corporation. In addition, the author was also partially supported by NSERC grant A 89858.

## BIBLIOGRAPHY

Aranda-Ordaz, F.J. (1981). On two families of transformations to additivity for binary regression data. *Biometrika* **68**, 357-363.

Balakrishnan, N. (ed) (1992). *Handbook of the Logistic Distribution*. Marcel Dekker, Inc. New York.

Balakrishnan, N. and Leung, M. Y. (1988). Order Statistics from the Type I Generalized Logistic Distribution. *Commun. Statist. - Simult. Comput.* **17**(1), 25-50.

Bickel, P.J. and Doksum, K.A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc* **76**, 296-311

Burr, I.W. (1942). Cumulative frequency functions. *Ann. Math Statist.* **13**, 215-232.

Copenhaver, T.W. and Mielke, P.W. (1977). Quantal analysis: A quantal assay refinement procedure. *Biometrics* **33**, 175-186.

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.

Czado, C. and Santner, T. J. (1992a). The Effect of Link Misspecification on Binary Regression Analysis. *J. Stat. Planning Inf.*,**33**, 213-231.

Czado, C. and Santner, T. J. (1992b). Orthogonalizing Link Transformation Families in Binary Regression Analysis, *Canad. J. Statist.* **20**, no 1, 51-61.

Czado, C. (1992a). On Selecting Parametric Link Transformation Families in Binary Regression Analysis, submitted.

Czado, C. (1992b). On Link Selection in Generalized Linear Models , *Advances in GLIM and Statistical Modelling*, L. Fahrmeir, B. Francis, R. Gilchrist, G. Tutz (eds), 60-65, Springer-Verlag, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) *J. Roy. Statist. Soc. (B)* **39** 1-38.

Duffy, D. E. and Santner, T. J. (1988). Estimating logistic regression probabilities, in *Statistical Decision Theory and Related Topics IV* **1**, S. S. Gupta, J. O. Berger (eds), 177-195, Springer-Verlag ,New York.

Duffy, D. E. and Santner, T. J. (1989). On the small sample properties of restricted maximum likelihood estimators for logistic regression models. *Comm. Statist. - Theor. Meth.* **18**, 959-989.

Guerrero, V. M. and Johnson, R. A. (1982). Use of Box-Cox transformation with binary response models. *Biometrika* **69**, 309-314.

Hinkley, D.V. and Runger, G. (1984). The analysis of transformed data. *J. Amer. Statist. Assoc.* **83**, 426-431.

Laird, N. M. (1978). Empirical Bayes Methods for two-way contingency tables. *Biometrika* **65**, 581-590.

Leonard, T. (1972) Bayesian methods for binomial data. *Biometrika* **59**, 581-589.

Leonard, T. (1975). Bayesian estimation methods for two-way contingency tables. *J. Roy. Statist. Soc. (B)* 23-27.

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

Prentice, R.L. (1975). Discrimination among some parametric models. *Biometrika* **64**, 607-614.

Prentice, R.L. (1976). A generalization of the probit and logistic methods for dose response curves. *Biometrics* **32**, 761-768.

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

Received September 1992; Revised March 1993