Hacettepe Journal of Mathematics and Statistics
Volume 33 (2004), 97 – 109

STOCHASTIC COVARIATES IN BINARY REGRESSION

Evrim Oral* and Süleyman Günay†

Received 27 : 04 : 2004 : Accepted 23 : 09 : 2004

Abstract

Binary regression has many medical applications. In applying the technique, the tradition is to assume the risk factor X as a non-stochastic variable. In most situations, however, X is stochastic. In this study, we discuss the case when X is stochastic in nature, which is more realistic from a practical point of view than X being non-stochastic. We show that our solutions are much more precise than those obtained by treating X as non-stochastic when, in fact, it is stochastic.

Keywords: Binary regression, Modified Maximum Likelihood Estimator, Robustness, Skew family, Symmetric family.

1. Introduction

A binary regression model typically is

$$\begin{aligned}\pi(x) = E(Y \mid X = x) &= \int_{-\infty}^{z} f(x)\,dx \\ &= F(z), \qquad (1.1)\end{aligned}$$

where $z = \gamma_0 + \gamma_1 x\,(\gamma_1 > 0)$, Y is a stochastic variable that assumes values 1 or 0 and X is a risk factor which in the literature has been treated as non-stochastic. In most situations, however, X is stochastic. Consider, for example, the following data:

(1) The 27 observations on (Y, X) given in Agresti [2, p. 88], where X measures the proliferative activity of cells after a patient receives an injection of tritiated thymidine and the response variable Y represents whether the patient achieves remission or not.

(2) The following 10 observations on (Y, X) given in Hosmer and Lemeshow [5, p.132]

$Y : 0, 1, 0, 0, 0, 0, 0, 1, 0, 1$

$X : 0.225, 0.487, -1.080, -0.870, -0.580, -0.640, 1.614, 0.352, -1.025, 0.929;$

*Department of Statistics, Middle East Technical University, Ankara 06531, Turkey.
†Department of Statistics, Hacettepe University, Ankara 06532, Turkey.

X is generated from normal $N(0,1)$. In the examples above, and many others (Aitkin *et al.*, [3]), X is clearly stochastic but has been treated as a non-stochastic variable. The function $f(x)$ has traditionally been taken to be logistic but Tiku and Vaughan [18] have extended the methodology to non-logistic density functions treating X as non-stochastic. The purpose here is to give solutions in the more realistic situations when X is stochastic and $f(x)$ is logistic or non-logistic. Since maximum likelihood methodology is intractable, modified likelihood methodology is invoked. The latter is known to yield MMLE (modified maximum likelihood estimators) as efficient as the MLE (maximum likelihood estimators). Unlike the MLE, the MMLE are explicit functions of sample observations and are easy to compute; see Vaughan [22] and Tiku and Vaughan [18]. In fact, Vaughan [23, p.228] states five very desirable properties of the MMLE. Moreover, as pointed out in Şenoğlu and Tiku [14, p.363], the MMLE are numerically the same (almost) as the MLE in all situations where authentic iterative MLE are available. See also Vaughan [23, p. 233]. The solutions we give are enormously more precise than those obtained by treating X as a non-stochastic variable when, in fact, it is stochastic.

2. Maximum Likelihood

Let Y be a binary random variable which assumes values $y_i = 1$ or 0 with probabilities θ and $1-\theta$, respectively, and let the corresponding observations on the risk factor X be denoted by $x_i\,(1 \le i \le n)$. The model is

$$F(z_i) = \pi(x_i) = E(Y|X = x_i) = \int_{-\infty}^{z_i} f(u)\,du, \tag{2.1}$$

where $z_i = \gamma_0 + \gamma_1 x_i$ and $f(u)$ is a completely specified density function; Y is presumed to increase with X so that γ_1 is a priori positive. Let $h(x)$ denote the probability density function of the stochastic variable X. The methodology developed here is applicable to any completely specified density f in (2.1) and any location-scale density $h(x)$. We consider, for illustration, two families of densities: (a) skew and (b) symmetric, which are prevalent in practice (Rasch [11]; Spjotvoll and Aastveit [12]; Tiku et al. [19]).

Skew family: Consider the family of Generalized Logistic densities

$$GL(b,\gamma_1) : h(x) = \frac{b\gamma_1 \mathrm{e}^{-(\gamma_0+\gamma_1 x)}}{\left(1+\mathrm{e}^{-(\gamma_0+\gamma_1 x)}\right)^{b+1}},\ -\infty < x < \infty. \tag{2.2}$$

Note that the probability density function of $Z = (X-\mu)/\sigma = \gamma_0 + \gamma_1 X$, ($\gamma_0 = -\mu/\sigma$, $\gamma_1 = 1/\sigma$), is

$$h(z) = \frac{be^{-z}}{\left(1+\mathrm{e}^{-z}\right)^{b+1}},\ -\infty < z < \infty. \tag{2.3}$$

The densities $f(u)$ and $h(z)$ are assumed to have the same functional forms although our methodology easily extends to situations where f and h are different from each other. The latter will perhaps be true if there are more than one risk factor and $z = \gamma_0 + \sum_i \gamma_1 x_i$.

For $b < 1$, (2.3) represents negatively skewed density functions. For $b > 1$, it represents positively skewed density functions. For $b = 1$, it is the well known logistic density. The mean and variance of (2.3) are

$$E(Z) = \psi(b) - (1) \text{ and } V(Z) = \psi'(b) + \psi'(1) \tag{2.4}$$

respectively The values of the psi-function $\psi(b)$ and its derivative $\psi'(b)$ are given in Tiku *et al.* [20, p. 1356]. See Abramowitz and Stegun [1] for analytical and computational aspects of psi-functions.

The likelihood function of the random sample (y_i, x_i), $1 \le i \le n$, is

$$L = L_X L_{Y|X} \propto \prod_{i=1}^{n} (\gamma_1 h(z_i)) \left\{ (F(z_i))^{y_i} (1 - F(z_i))^{1-y_i} \right\}. \tag{2.5}$$

This gives

$$\ln L \propto n \ln \gamma_1 + \sum_{i=1}^{n} \{\ln h(z_i) + y_i \ln F(z_i) + (1 - y_i) \ln (1 - F(z_i))\}; \tag{2.6}$$

$z_i = \gamma_0 + \gamma_1 x_i$ and $F(z_i) = \left(1 + \mathrm{e}^{-z_i}\right)^{-b}$.

The likelihood equations are expressions in terms of the intractable functions

$$g(z) = \mathrm{e}^{-z} / \left(1 + \mathrm{e}^{-z}\right), \; g_1(z) = f(z) / F(z) \text{ and } g_2(z) = f(z) / \{1 - F(z)\}. \tag{2.7}$$

They have no explicit solutions and are almost impossible to solve by iterative methods. The MLE are, therefore, elusive. See also Tiku and Vaughan [18] who work with the conditional likelihood function $L_{Y|X}$ as do other authors (Agresti [2]; Hosmer and Lemeshow [5]; Kleinbaum [6]).

To obtain the MMLE, we first express the likelihood equations $\partial \ln L / \partial \gamma_0 = 0$ and $\partial \ln L / \partial \gamma_1 = 0$ in terms of the ordered variates $z_{(i)}$. Since γ_1 is a priori positive,

$$z_{(i)} = \gamma_0 + \gamma_1 x_{(i)}, 1 \le i \le n, \tag{2.8}$$

where $x_{(i)}$ are the order statistics of the random sample x_i, $(1 \le i \le n)$. Since complete sums are invariant under ordering, e.g. $\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} z_{(i)}$, we have

$$\frac{\partial \ln L}{\partial \gamma_0} = -n + \sum_{i=1}^{n} \left\{ (b+1) g\left(z_{(i)}\right) + w_i g_1\left(z_{(i)}\right) - (1 - w_i) g_2\left(z_{(i)}\right) \right\} = 0 \tag{2.9}$$

and

$$\begin{aligned} \frac{\partial \ln L}{\partial \gamma_1} &= \frac{n}{\gamma_1} - \sum_{i=1}^{n} x_{(i)} + \sum_{i=1}^{n} x_{(i)} \left\{ (b+1) g\left(z_{(i)}\right) + w_i g_1\left(z_{(i)}\right) - (1 - w_i) g_2\left(z_{(i)}\right) \right\} \\ &= 0, \end{aligned} \tag{2.10}$$

where $w_i = y_{[i]}$ is the concomitant of $x_{(i)}$, i.e., $y_{[i]}$ is that observation y_i which is coupled with $x_{(i)}$ when (y_i, x_i) are ordered with respect to x_i, $(1 \le i \le n)$. As mentioned earlier, (2.8)-(2.9) are almost impossible to solve.

3. Modified Likelihood

To obtain the modified likelihood equations, we linearize the functions $g(z)$, $g_1(z)$ and $g_2(z)$ by using the first two terms of their Taylor series expansions as follows:

$$\begin{aligned} g\left(z_{(i)}\right) &\cong g\left(t^*_{(i)}\right) + \left(z_{(i)} - t^*_{(i)}\right) \left\{ \frac{d}{dz} g(z) \right\}_{z = t^*_{(i)}} \\ &= \alpha_i - \beta_i z_{(i)}, \; 1 \le i \le n, \end{aligned} \tag{3.1}$$

where

$$\alpha_i = (1 + \mathrm{e}^{\alpha_i})^{-1} + \beta_i a_i \text{ and } \beta_i = \mathrm{e}^{a_i} / (1 + \mathrm{e}^{a_i})^2, \; a_i = t^*_{(i)}; \tag{3.2}$$

a_i is determined by the equation

$$\int_{-\infty}^{a_i} h(z) \, dz = \frac{i}{n+1}, \; 1 \le i \le n. \tag{3.3}$$

Similarly,

$$g_1\left(z_{(i)}\right) \cong \alpha_{1i} - \beta_{1i} z_{(i)} \text{ and } g_2\left(z_{(i)}\right) = \alpha_{2i} + \beta_{2i} z_{(i)}, \; 1 \le i \le n, \tag{3.4}$$

where

$$\beta_{1i} = \left\{ f^2\left(t_{(i)}\right) - F\left(t_{(i)}\right) f'\left(t_{(i)}\right) \right\} / F^2\left(t_{(i)}\right) \text{ and } \alpha_{1i} = g_1\left(t_{(i)}\right) + \beta_{1i} t_{(i)}, \tag{3.5}$$

and

$$\begin{aligned} &\beta_{2i} = \left\{ f^2\left(t_{(i)}\right) + \left(1 - F\left(t_{(i)}\right)\right) f'\left(t_{(i)}\right) \right\} / \left(1 - F\left(t_{(i)}\right)\right)^2, \\ &\alpha_{2i} = g_2\left(t_{(i)}\right) - \beta_{2i} t_{(i)}. \end{aligned} \tag{3.6}$$

Here, $t_{(i)}$ is determined by the equation

$$\int_{-\infty}^{t_{(i)}} f(z)\, dz = \frac{i}{n+1},\ 1 \le i \le n. \tag{3.7}$$

For the Generalized Logistic,

$$t_{(i)} = -\ln\left\{ q_i^{-\frac{1}{b}} - 1 \right\},\ q_i = i/(n+1). \tag{3.8}$$

Since $f(u)$ and $h(z)$ are assumed to have the same forms, $a_{(i)} = t_{(i)}$ $(1 \le i \le n)$.

Incorporating (3.1) and (3.4) in (2.9)-(2.10) gives the modified likelihood equations which have the following beautiful expressions:

$$\frac{\partial \ln L}{\partial \gamma_0} \cong \frac{\partial ln L^*}{\partial \gamma_0} = -n + \sum_{i=1}^{n} \left\{ \delta_i - m_i z_{(i)} \right\} = 0 \tag{3.9}$$

and

$$\frac{\partial \ln L}{\partial \gamma_1} \cong \frac{\partial ln L^*}{\partial \gamma_1} = \frac{n}{\gamma_1} - \sum_{i=1}^{n} x_{(i)} + \sum_{i=1}^{n} x_{(i)} \left\{ \delta_i - m_i z_{(i)} \right\} = 0 \tag{3.10}$$

where

$$\delta_i = w_i \alpha_{1i} - (1 - w_i)\alpha_{2i} + (b+1)\alpha_i \text{ and } m_i = w_i \beta_{1i} + (1 - w_i)\beta_{2i} + (b+1)\beta_i. \tag{3.11}$$

The solutions of (3.9)-(3.10) are the MMLE $\hat{\gamma_0}$ and $\hat{\gamma_1}$ of γ_0 and γ_1, respectively:

$$\hat{\gamma_0} = \{(\delta - n)/m\} - \hat{\gamma_1}\overline{x}_{(.)} \text{ and } \hat{\gamma_1} = \left\{ B + \sqrt{(B^2 + 4nC)} \right\} / 2C \tag{3.12}$$

where

$$\begin{aligned} &\delta = \sum_{i=1}^{n} \delta_i,\ m = \sum_{i=1}^{n} m_i,\ \overline{x}_{(.)} = \frac{1}{m} \sum_{i=1}^{n} m_i x_{(i)}, \\ &B = \sum_{i=1}^{n} (\delta_i - 1)\left(x_{(i)} - \overline{x}_{(.)}\right), \text{ and} \\ &C = \sum_{i=1}^{n} m_i \left(x_{(i)} - \overline{x}_{(.)}\right)^2 = \sum_{i=1}^{n} m_i x_{(i)}^2 - \frac{1}{m}\left(\sum_{i=1}^{n} m_i x_{(i)} \right)^2. \end{aligned} \tag{3.13}$$

Revised estimates: Following Lee *et al.* [7] and Tiku and Vaughan [18, p. 888], in order to sharpen the MMLE, we calculate the coefficients $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ from (3.5)-(3.6) by replacing $t_{(i)}$ by

$$t_i = \hat{\gamma_0} + \hat{\gamma_1} x_{(i)},\ 1 \le i \le n, \tag{3.14}$$

and calculate the revised estimates from (3.12)-(3.13). We may repeat this process a few times until the estimates stabilize to, say, three decimal places. No revision is needed in the coefficients (α_i, β_i) in (3.1); they are computed from (3.2) once and for all. See also Tiku and Suresh [17] and Vaughan [22, 23].

4. Asymptotic Variances and Covariances

It has been rigourously proved by Vaughan and Tiku [24] that the MMLE are asymptotically equivalent to the MLE. Bhattacharyya [4] establishes this result for censored samples. Therefore, the MMLE $\hat{\gamma}_0$ and $\hat{\gamma}_1$ above are asymptotically unbiased and efficient. Their asymptotic variance-covariance matrix $\mathbf{V}$ is given by $\mathbf{I}^{-1}(\gamma_0, \gamma_1)$, where $\mathbf{I}$ is the Fisher information matrix consisting of the following elements:

$$I_{11} = -E\left(\frac{\partial^2 \ln L}{\partial \gamma_0^2}\right) = Q + P_1, \tag{4.1}$$

$$I_{12} = -E\left(\frac{\partial^2 \ln L}{\partial \gamma_0 \partial \gamma_1}\right)$$
$$= \frac{1}{\gamma_1}\left\{-\left(Q + P_1\right)\gamma_0 + \left[\psi\left(b\right) - \psi\left(1\right)\right]Q + \left[\psi\left(b+1\right) - \psi\left(2\right)\right]P_1\right\} \tag{4.2}$$

and

$$I_{22} = -E\left(\frac{\partial^2 \ln L}{\partial \gamma_1^2}\right)$$
$$= \frac{1}{\gamma_1^2}((Q + P_1)\gamma_0^2 - 2\gamma_0\left(\left[\psi\left(b\right) - \psi\left(1\right)\right]Q + \left[\psi\left(b+1\right) - \psi\left(2\right)\right]P_1\right) +$$
$$+ \left(\left[\psi\left(b\right) - \psi\left(1\right)\right]^2 + \psi'\left(b\right) + \psi'\left(1\right)\right)Q +$$
$$+ \left(\left[\psi\left(b+1\right) - \psi\left(2\right)\right]^2 + \psi'\left(b+1\right) + \psi'\left(2\right)\right)P_1 + n), \tag{4.3}$$

where

$$Q_i = \frac{f^2\left(z_i\right)}{F\left(z_i\right)\left(1 - F\left(z_i\right)\right)},\ Q = \sum_{i=1}^{n} Q_i \text{ and } P_1 = \frac{nb}{b+2}. \tag{4.4}$$

The expressions (4.1)-(4.3) are obtained along the same lines as in Tiku and Vaughan [18, Section 5] and realizing that for a bounded bivariate random function $g(Z, Y)$,

$$E\left\{g\left(Z, Y\right)\right\} = E_Z\left\{E_{Y/Z}\, g\left(Z, Y\right)\right\}. \tag{4.5}$$

Thus,

$$\mathbf{V} = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix}^{-1}. \tag{4.6}$$

In particular, the asymptotic variance of $\hat{\gamma}_1$ is given by

$$\operatorname{var}\left(\hat{\gamma}_1/\gamma_1\right) \cong \left(Q + P_1\right)/\Delta, \tag{4.7}$$

where

$$\Delta = (\left[\psi'\left(b\right) - \psi'\left(1\right)\right]Q\left(Q + P_1\right) + \left[\psi'\left(b+1\right) + \psi'\left(2\right)\right]P_1\left(Q + P_1\right) +$$
$$+ (\left[\psi\left(b\right) - \psi\left(1\right)\right] - \left[\psi\left(b+1\right) - \psi\left(2\right)\right])^2 QP_1 + n\left(Q + P_1\right),$$

which is free of γ_0.

An estimate of the variance of $\hat{\gamma}_1$ is obtained by replacing z_i by $\hat{z}_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_i$ in (4.4). The standard error of $\hat{\gamma}_1$ is the square root of this estimated variance and, similarly, for $\hat{\gamma}_0$.

For $b = 1$ (logistic density), (4.7) simplifies and since $\psi'(1) = 1.6449$ and

$$\psi'\left(2\right) = 1.6449,\ \operatorname{var}\left(\hat{\gamma}_1/\gamma_1\right) \cong 1/\left(3.2898Q + 1.2898P_1 + n\right). \tag{4.8}$$

We now give a few examples to illustrate the enormous gain in efficiency of the estimators when the complete likelihood function L is used as against using only the conditional likelihood $L_{Y|X}$.

4.1. Example. Consider the widely reported CHD (coronary heart disease) data on 100 randomly chosen patients (Hosmer and Lemeshow [5]). Here, X represents the age and Y the presence or absence of CHD. The density $f(u)$ in (2.1) has traditionally been taken to be logistic. With logistic $h(x)$, i.e. $GL(1, \gamma_1)$ in (2.2), we have the MMLE and their standard errors reported in Table 1. Only two iterations were needed for the estimates to stabilize to three decimal places. Also reported are the MLE and the MMLE (and their standard errors) based only on the conditional likelihood $L_{Y|X}$, reproduced from Tiku and Vaughan [18, p. 890]. It can be seen that the MMLE based on the complete likelihood (2.5) are enormously more precise, that is, they have considerably smaller standard errors than those based only on the conditional likelihood.

Table 1. Estimates and their Standard Errors for the CHD Data, $n = 100$**.**

		Coefficient	Estimate	Standard Error
Conditional Likelihood	ML	γ_0	-5.310	1.134
		γ_1	0.111	0.024
	MML	γ_0	-5.309	1.134
		γ_1	0.111	0.024
Complete Likelihood	MML	γ_0	-6.181	0.463
		γ_1	0.136	0.010

It may be noted that the MMLE based on the complete likelihood are not much different numerically from those based only on the conditional likelihood.

4.2. Example. Consider the Agresti [2, p.88] data (27 observations) mentioned earlier. The MLE and the MMLE and their standard errors are given in Table 2. Again, the MMLE based on the complete likelihood function are not very different numerically from those based on the conditional likelihood but are enormously more precise.

Table 2: Estimates and their Standard Errors for Agresti's Data, $n = 27$**.**

		Coefficient	Estimate	Standard Error
Conditional Likelihood	ML	γ_0	-3.777	$*$
		γ_1	0.145	0.059
	MML	γ_0	-3.777	1.379
		γ_1	0.145	0.059
Complete Likelihood	MML	γ_0	-3.688	0.576
		γ_1	0.175	0.024

$*$: Not given in Agresti [2]

Simulation study: To illustrate further the gain in efficiency which the MMLE (3.12) provide over those based only on the conditional likelihood (Tiku and Vaughan [18]), i.e.,

$$(4.9) \qquad \hat{\gamma}_{1c} = \sum_{i=1}^{n} \delta_i \left(x_{(i)} - \overline{x}_a\right) / \sum_{i=1}^{n} m_i \left(x_{(i)} - \overline{x}_a\right)^2 \text{ and } \hat{\gamma}_{0c} = (\delta/m) - \hat{\gamma}_{1c}\overline{x}_a,$$

$$\delta = \sum_i \delta_i, \; m = \sum_i m_i, \; \delta_i = \alpha_{1i} w_i - \alpha_{2i}(1 - w_i),$$

$$m_i = \beta_{1i} w_i + \beta_{2i}(1 - w_i) \text{ and } \overline{x}_a = (1/m) \sum_i m_i x_{(i)}.$$

we did an extensive Monte Carlo study as follows:

We generated a random sample $x_1, x_2, \ldots, x_n$ of size n from (2.2) and calculated the MMLE of γ_0 and γ_1 (Tiku and Suresh [17]; Vaughan [22]) from the order statistics

$$x_{(i)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

Denote them by $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$. Specifically,

$$\begin{aligned} \hat{\gamma}_{00} &= \frac{1}{m}\sum_{i=1}^{n}\left(\alpha_i - \frac{1}{b+1}\right) - \hat{\gamma}_{10}\overline{x}_{(..)} \text{ and },\\ \hat{\gamma}_{10} &= \left\{-B_0 + \sqrt{B_0^2 + 4nC_0}\right\}/2C_0, \end{aligned} \tag{4.10}$$

where

$$\begin{aligned} m &= \sum_{i=1}^{n}\beta_i,\ \overline{x}_{(..)} = (1/m)\sum_{i=1}^{n}\beta_i x_{(i)},\\ B_0 &= (b+1)\sum_{i=1}^{n}\left(\alpha_i - \frac{1}{b+1}\right)\left(x_{(i)} - \overline{x}_{(..)}\right) \text{ and}\\ C_0 &= (b+1)\sum_{i=1}^{n}\beta_i\left(x_{(i)} - \overline{x}_{(..)}\right)^2; \end{aligned} \tag{4.11}$$

the values of α_i and β_i being obtained from (3.2).

We generated the concomitant binary observations $w_i = y_{[i]}$, $(1 \leq i \leq n)$, by calculating the probability

$$\hat{P}_i = \left(1 + \mathrm{e}^{-\hat{z}_{(i)}}\right)^{-b},\ \hat{z}_{(i)} = \hat{\gamma}_{00} + \hat{\gamma}_{10}x_{(i)} \tag{4.12}$$

and defining

$$w_i = \begin{cases} 1 & \text{if } U_i \leq \hat{P}_i,\\ 0 & \text{otherwise,} \end{cases} \tag{4.13}$$

where U_i, $(1 \leq i \leq n)$, are independent uniform (0,1) variates.

The observations $\left(w_i, x_{(i)}\right)$, $1 \leq i \leq n$, so generated were substituted in (3.12) and the MMLE $\hat{\gamma}_0$ and $\hat{\gamma}_1$ obtained. The corresponding conditional likelihood MMLE were obtained from the equations (4.9). Both sets of estimates were the result of three iterations. The procedure was repeated $N = [100000/n]$ (integer value) times. The means and variances of the resulting estimates were calculated for $b = 0.5, 1, 2$ and 4. Both were found to be (almost) unbiased. The MMLE $\hat{\gamma}_0$ and $\hat{\gamma}_1$, however, were found to be enormously more efficient than the conditional MMLE for all values of b. For illustration, we reproduce in Table 3 the means and variances of the estimators for $b = 1$. It can be seen that the MMLE based on the complete likelihood are very much more efficient than those based only on the conditional likelihood.

It is interesting to see that the mean and variance of $\hat{\gamma}_1/\gamma_1$ are almost invariant to γ_0 for all n, in agreement with the fact that $\hat{\gamma}_1$ is (almost) unbiased for γ_1 and the equation (4.7) is free of γ_0. Incidentally, the simulated variance of $\hat{\gamma}_1/\gamma_1$ is close to those obtained from (4.7) for large n (> 100)

Table 3. Simulated Means and Variances: (a) Mean and (b) Variance.

			Complete Likelihood				Conditional Likelihood			
			(a)		(b)		(a)		(b)	
γ_0	γ_1	n	$\hat{\gamma}_0$	$\hat{\gamma}_1/\gamma_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1/\gamma_1$	$\hat{\gamma}_{(0)}$	$\hat{\gamma}_{(1)}/\gamma_1$	$\hat{\gamma}_{(0)}$	$\hat{\gamma}_{(1)}/\gamma_1$
0	0.001	50	0.001	1.018	0.075	0.017	0.003	1.112	0.233	0.155
		100	-0.005	1.007	0.035	0.008	-0.011	1.062	0.101	0.070
	0.10	50	0.006	1.023	0.079	0.017	0.001	1.121	0.229	0.168
		100	0.003	1.009	0.039	0.008	-0.000	1.048	0.108	0.061
	1.00	50	-0.005	1.022	0.075	0.017	-0.010	1.132	0.223	0.165
		100	-0.005	1.011	0.037	0.008	-0.011	1.066	0.103	0.069
2	0.001	50	2.029	1.017	0.142	0.017	2.245	1.127	0.892	0.161
		100	2.025	1.011	0.073	0.008	2.124	1.056	0.350	0.063
	0.10	50	2.035	1.016	0.139	0.016	2.239	1.119	0.866	0.159
		100	2.018	1.005	0.070	0.008	2.111	1.048	0.375	0.067
	1.00	50	2.048	1.024	0.136	0.015	2.278	1.139	0.891	0.170
		100	2.021	1.008	0.070	0.008	2.114	1.055	0.352	0.059

4.3. Remark. The assumed values of γ_0 and γ_1 may as well be used in (4.12) to generate w_i from (4.13). That does not change the values in Table 3 in any substantial way.

5. Hypothesis Testing

Since X is a genuine risk factor, its effect on Y will logically be never zero. The real issue, therefore, is whether it has some effect howsoever small. Testing $H_0 : \gamma_1 = \gamma_{10}\,(> 0)$ is, therefore, of major importance. To test H_0, we propose the statistic

(5.1) $$W = \hat{\gamma}_1/\gamma_{10}.$$

Large values of W lead to the rejection of H_0 in favour of $H_1 : \gamma_1 > \gamma_{10}$. Since $\hat{\gamma}_1$ is asymptotically equivalent to the MLE, the asymptotic null distribution of W is normal with mean 1 and variance given by the right hand side of (4.7). Simulations reveal, however, that it takes a large sample size n (> 100) to attain near-normality of the null distribution of W. To study the null distribution of W for small n, we simulated the coefficients of skewness $\beta_1^* = \mu_3^2/\mu_2^3$ and kurtosis $\beta_2^* = \mu_4/\mu_2^2$ of W. Interestingly, $\mu_3 > 0$ and β_1^* and β_2^* are close to the Type III line in the Pearson plane (Pearson [9]; Tiku [15,16]; Pearson and Tiku [10, Fig.1]), i.e.,

(5.2) $$|\beta_2^* - (3 + 1.5\beta_1^*)| \le 0.5.$$

For $n = 20$, $\gamma_0 = 2$ and $\gamma_1 = 0.001$, for example, $\beta_1^* = 0.732$ and $\beta_2^* = 4.309$ so that (5.2) equals 0.211; for $n = 100$, $\beta_1^* = 0.139$ and $\beta_2^* = 3.153$, and (5.2) equals 0.056. The Pearson-Tiku 3-moment chi-square approximation is applicable. This works as follows: Let

(5.3) $$\chi^2 = (W + c)/d,$$

where χ^2 is a chi-square variate with ν degrees of freedom. Determine ν, d and c such that the first three moments on both sides agree. This gives

(5.4) $$\nu = 8/\beta_1^*,\ d = \sqrt{(\mu_2/2\nu)} \text{ and } c = b\nu - \mu_1'.$$

Here μ_1' and μ_2 are the simulated mean and variance of W. Thus, the $100\,(1-\alpha)\,\%$ point of W is given by

(5.5) $\quad W_\alpha = d\chi^2_\alpha\,(\nu) - c,$

where $\chi^2_\alpha\,(\nu)$ is the $100\,(1-\alpha)\,\%$ point of a chi-square distribution with ν degrees of freedom. The 3-moment chi-square approximation gives remarkably accurate values. For example, we have the following simulated values of the probability $P\,(W \geq W_\alpha | H_0)$ with its value presumed to be 0.050:

Simulated Values of The Probability

n	$b = 0.5$	1	2	4
20	0.049	0.044	0.046	0.048
50	0.051	0.050	0.049	0.047
100	0.049	0.045	0.051	0.051

5.1. Remark. Since $\hat{\gamma}_1$ is as efficient as the MLE, it will not be easy to improve over the W test so far as its power is concerned.

Since X has been assumed to be non-stochastic, it is now possible to study the robustness of the W test as follows.

Robustness: In practice, a value of the shape parameter b in (2.3) is located with the help of $Q-Q$ plots and/or formal goodness-of-fit tests; see specifically Tiku and Vaughan [18, Appendix B]. In spite of ones best efforts, however, it might not be possible to locate the exact value of b. Moreover, the data might contain outliers or be contaminated. That brings the robustness issue in focus. Assume, for illustration, that the true value in (2.3) is $b = 1$ (logistic distribution) which we will call the population model. As plausible alternatives, we consider the following which we will call sample models; $\sigma = 1/\gamma_1$.

Misspecified model:

Student's t distribution with degrees of freedom

(1) $\upsilon = 9$,
(2) $\upsilon = 19$.

Outlier model:

(3) $(n-r)$ observations come from $GL\,(1,\sigma)$ and r (we do not know which) come from $GL\,(1, 4\sigma)$, $r = [0.5 + 0.1n]$.
(4) $(n-r)$ observations come from $GL\,(0.5,\sigma)$ and r come from $GL\,(0.5, 4\sigma)$.

Mixture model:

(5) $0.90GL\,(1,\sigma) + 0.10GL\,(1, 4\sigma)$,
(6) $0.90GL\,(0.5,\sigma) + 0.10GL\,(0.5, 4\sigma)$.

Contamination model:

(7) $(n-r)$ observations come from $GL\,(1,\sigma)$ and r from uniform $U\,(0,1)$.

Models (4) and (6) represent skew distributions. All other models represents symmetric distributions.

To study the robustness of Type I error of the W test above, since the assumed population model is the logistic $GL\,(1,\sigma)$, we take $b = 1$ in all the equations above and use the corresponding W_α in (5.5) for all the alternative models (1)–(7). The simulated Type I errors are given in Table 4. It can be seen that the W test is remarkably robust. This is typical of the hypothesis testing procedures based on the MMLE since the coefficients β_i have umbrella or half-umbrella ordering; see specifically Şenoğlu and Tiku [13,14].

Table 4. Simulated Values of The Type I Error; $\gamma_0 = 0$, $\gamma_1 = 0.001$.

Model	$n = 50$	60	80	100	Model	$n = 50$	60	80	100
Logistic	0.050	0.052	0.052	0.048	(1)	0.049	0.051	0.048	0.048
(2)	0.049	0.047	0.048	0.051	(3)	0.049	0.049	0.048	0.052
(4)	0.053	0.050	0.050	0.050	(5)	0.051	0.047	0.046	0.050
(6)	0.051	0.049	0.050	0.051	(7)	0.050	0.049	0.050	0.052

The random numbers generated from all the models in the table were standardized to have variance 1. Although we have reported values only for $\gamma_0 = 0$ and $\gamma_1 = 0.001$, the results are essentially the same for other values of γ_0 and γ_1. Similarly, the power of the test is not diminished in any substantial way under plausible deviations from an assumed model. We omit details for conciseness. Thus, the W test has both criterion robustness as well as efficiency robustness.

6. Symmetric family

Consider the situation when X has density

$$h(x) = \frac{\gamma_1}{\sqrt{k}B\left(\frac{1}{2}, p - \frac{1}{2}\right)} \left\{1 + \frac{1}{k}(\gamma_0 + \gamma_1 x)^2\right\}^{-p}, \quad -\infty < x < \infty; \tag{6.1}$$

$k = 2p - 3$, $p \geq 2$. The probability density function of $Z = \gamma_0 + \gamma_1 X$ is

$$h(z) = \frac{1}{\sqrt{k}B\left(\frac{1}{2}, p - \frac{1}{2}\right)} \left\{1 + \frac{1}{k}z^2\right\}^{-p}, \quad -\infty < z < \infty. \tag{6.2}$$

Realize that $t = \sqrt{(\nu/k)}Z$ has the Student's t distribution with $\nu = 2p - 1$ degrees of freedom.

Given an ordered sample $\left(w_i, x_{(i)}\right)$, $1 \leq i \leq n$, the likelihood equations are $(w_i = y_{[i]})$,

$$\frac{\partial lnL}{\partial \gamma_0} = \sum_{i=1}^{n} \left\{-\frac{2p}{k} g\left(z_{(i)}\right) + w_i g_1\left(z_{(i)}\right) - (1 - w_i)\, g_2\left(z_{(i)}\right)\right\} = 0 \tag{6.3}$$

and

$$\frac{\partial lnL}{\partial \gamma_1} = \frac{n}{\gamma_1} + \sum_{i=1}^{n} x_{(i)} \left\{-\frac{2p}{k} g\left(z_{(i)}\right) + w_i g_1\left(z_{(i)}\right) - (1 - w_i)\, g_2\left(z_{(i)}\right)\right\} = 0 \tag{6.4}$$

where

$$g(z) = z/\left\{1 + (1/k)\, z^2\right\}, \ g_1(z) = f(z)/F(z) \text{ and } g_2(z) = f(z)/\{1 - F(z)\}. \tag{6.5}$$

Here, $f(u)$ and $h(z)$ are assumed to be the same functions (6.2) and $F(z) = \int_{-\infty}^{z} f(u)\, du$. Again, it is almost impossible to solve the equations (6.3)–(6.4).

Proceeding exactly along the same lines as in Section 2, modified likelihood equations are obtained. The solutions of these equations are the following MMLE:

$$\hat{\gamma}_0 = (\delta/m) - \hat{\gamma}_1 \overline{x}_{(.)} \text{ and } \hat{\gamma}_1 = \left\{B + \sqrt{B^2 + 4nC}\right\}/2C, \tag{6.6}$$

where

$$B = \sum_{i=1}^{n} \delta_i \left(x_{(i)} - \overline{x}_{(.)} \right) \text{ and}$$

$$(6.7) \quad C = \sum_{i=1}^{n} m_i \left(x_{(i)} - \overline{x}_{(.)} \right)^2 = \sum_{i=1}^{n} m_i x_{(i)}^2 - (1/m) \left(\sum_{i=1}^{n} m_i x_{(i)} \right)^2 ;$$

$$\delta_i = w_i \alpha_{1i} - (1 - w_i) \alpha_{2i} - (2p/k) \alpha_i \text{ and } m_i = w_i \beta_{1i} + (1 - w_i) \beta_{2i} + (2p/k) \beta_i .$$

The coefficients α_i and β_i are given by

$$(6.8) \quad \alpha_i = \frac{(2/k) a_i^3}{\{1 + (1/k) a_i^2\}^2} \text{ and } \beta_i = \frac{1 - (1/k) a_i^2}{\{1 + (1/k) a_i^2\}^2},$$

where a_i is determined from the equation

$$(6.9) \quad F(a_i) = \int_{-\infty}^{a_i} h(z)\, dz = \frac{i}{n+1}, \quad 1 \le i \le n.$$

An IMSL subroutine is available to evaluate (6.9) which is essentially the cumulative density function of the Student's t distribution.

The coefficients $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ are given in (3.5)–(3.6) with $f(z)$ replaced by the density on the right hand side of (6.2), and $F(z) = \int_{-\infty}^{z} f(u)\, du$.

6.1. Remark. If $\beta_1 > 0$, then all the coefficients β_i are positive since they increase until the middle value and then decrease in a symmetric fashion (umbrella ordering). If, however, β_1 assumes a negative value (which happens only if p is small and n is large), $\hat{\sigma}$ might cease to be real and positive. In such a situation, α_i and β_i in (6.7) are replaced by $\alpha_i^* = 0$ and $\beta_i^* = \left\{1 + (1/k) a_i^2\right\}$ respectively. This ensures that $\hat{\gamma}_1$ is always real and positive. The asymptotic efficiency of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ is not affected since $\alpha_i + \beta_i z_{(i)} \cong \alpha_i^* + \beta_i^* z_{(i)}$ realizing that $z_{(i)} - t_{(i)} \cong 0$ (asymptotically). See also Tiku et al. [21].

Information matrix: Because of the symmetry of (6.1), the elements of the Fisher information matrix are simpler than those in (4.1)–(4.3) and are given by

$$(6.10) \quad \begin{aligned} &I_{11} = Q + P,\ I_{12} = -(\gamma_0/\gamma_1)(Q + P) \text{ and} \\ &I_{22} = \left(1/\gamma_1^2\right) \left\{ (Q + P) \gamma_0^2 + (Q + R) \right\}. \end{aligned}$$

Here,

$$(6.11) \quad P = np(p - 1/2)/(p + 1)(p - 3/2) \text{ and } R = 2n(p - 1/2)/(p + 1),\ (p \ge 2),$$

and Q_i, Q have exactly the same expressions as (4.4).

In particular, for the asymptotic variance

$$(6.12) \quad var(\hat{\gamma}_1/\gamma_1) \cong 1/(Q + R),$$

which is free of γ_0.

Efficiency: We simulated the means and variances of the MMLE $\hat{\gamma}_0$ and $\hat{\gamma}_1$ and compared them with the corresponding estimators (Tiku and Vaughan [18]) based only on the conditional likelihood function $L_{Y|X}$. As for the $GL(b, \gamma_1)$ family (2.2), $\hat{\gamma}_0$ and $\hat{\gamma}_1$ were found to be (almost) unbiased and enormously more efficient. The test of $H_0 : \gamma_1 = \gamma_{10}$ based on the statistic $W = \hat{\gamma}_1/\gamma_{10}$ was found to be remarkably robust to plausible deviations from an assumed value of p in (6.1) and to data anomalies, e.g., outliers and contaminated data. Details are given in Oral [8]. To save space we do not reproduce them here.

Generalization: The methodology above readily extends to the situation when

$$z = \gamma_0 + \sum_{i=1}^{k} \gamma_i x_i$$

and the risk factors X_i $(1 \le i \le k)$ are independently distributed with densities $h_i(x)$. The coefficients γ_i are a priori all positive. The MMLE based on the conditional likelihood

$$L_{Y|X_1,X_2,\ldots,X_k}$$

is given in Tiku and Vaughan [18]. It will be of great interest to extend the methodology to correlated risk factors. Extensions to censored (Type I and Type II) data are also of enormous interest from a practical point of view. Another interesting extension is to the situation when Y is multinomial and assumes more than two values.

In conclusion it must be said that it is advantageous to use the complete likelihood function to obtain efficient and robust estimators. Conditional likelihood was perhaps used to make estimation computationally feasible. Modified likelihood methodology makes estimation easy both analytically and computationally. Its use with the complete likelihood function gives estimators which are enormously more efficient than those based only on the conditional likelihood function. Moreover, the method is applicable to any location-scale distribution and no specialized computer software is needed to compute the MMLE.

Acknowledgment. We would like to thank to Professor M. L. Tiku for supervising this research.

References

[1] Abramowitz, M. and Stegun, I. A. *Handbook of Mathematical Functions*, (Dover, New York, 1965).

[2] Agresti, A. *Categorical Data Analysis*, (John Wiley and Sons, New York, 1996).

[3] Aitkin, M., Anderson, D., Francis, B. and Hinde, J. *Statistical Modelling In GLIM*, (Oxford Science, New York, 1989).

[4] Bhattacharyya, G. K. *The asymptotics of maximum likelihood and related estimators based on Type II censored data*, J. American Statistical Association **80**, 398–404, 1985.

[5] Hosmer, D. W. and Lemeshow, S. *Applied Logistic Regression*, (John Wiley and Sons, New York, 1989).

[6] Kleinbaum, D. G. *Logistic Regression*, (Springer, New York, 1994).

[7] Lee, K. R., Kapadia, C. H. and Dwight, B. B. *On estimating the scale parameter of Rayleigh distribution from censored samples*, Statistische Hefte **21**, 14–20, 1980.

[8] Oral, E. *Stochastic and Nonstochastic Covariates in Binary Regression*, Unpublished Ph. D. Thesis, (Hacettepe Univ., Ankara, Turkey., 2002).

[9] Pearson, E. S. *Some problems arising in approximating the probability distributions using moments*, Biometrika **50**, 95–112, 1963.

[10] Pearson, E. S. and Tiku, M. L. *Some notes on the relationship between the distributions of central and non-central F*, Biometrika **57**, 175–179, 1970.

[11] Rasch, D. *Probleme der Angewandten Statistik, Robustheit I: Fz. Für Tierproduktion*, Dummerstorf-Rostock, Heft 4, 1980.

[12] Spjotvoll, E. and Aastveit, A. H. *Comparison of Robust Estimators on Some Data from Field Experiments*, Scand. J. Statist. **7**, 1–13, 1980.

[13] Senoglu, B. and Tiku, M. L. *Analysis of variance in experimental design with nonnormal rrror distributions*, Communications in Statistics-Theory and Methods **30**, 1335–1352, 2001.

[14] Senoglu, B. and Tiku, M. L. *Linear contrasts in experimental design with nonidentical error distributions*, Biometrical J. **44**, 359–374, 2002.

[15] Tiku, M. L. *Laguerre series forms of non-central Chi-square and F distributions*, Biometrika **52**, 415–427, 1965.

[16] Tiku, M. L. *Distribution of the derivative of the likelihood function*, Nature **210**, 766, 1966.

[17] Tiku, M. L. and Suresh, R. P. *A new method of estimation for location and scale parameters*, Journal of Statistical Planning and Inference **30**, 281–292, 1992.

[18] Tiku, M. L. and Vaughan, D. C. *Logistic and nonlogistic density functions in binary regression with nonstochastic covariates*, Biometrical J. 39, 883–898, 1997.

[19] Tiku, M. L., Tan, W. Y. and Balakrishnan, N. *Robust Inference*, (Marcel Dekker, New York, 1986).

[20] Tiku, M. L., Wong, W. K. and Bian, G. *Time series models with asymmetric innovations*, Communications in Statistics-Theory and Methods **28**, 1331–1360, 1999.

[21] Tiku, M. L., Wong, W. K. and Vaughan, D. C. *Time series models in non-normal situations: Symmetric innovations*, J. Time Series Analysis **21**, 571–596, 2000.

[22] Vaughan, D. C. *On the Tiku-Suresh method of estimation*, Communications in Statistics-Theory and Methods **21**, 451–469, 1992.

[23] Vaughan, D. C. *The generalized secant hyperbolic distribution and its properties*, Communications in Statistics-Theory and Methods **31**, 219–238, 2002.

[24] Vaughan, D. C. and Tiku M., L. *Estimation and hypothesis testing for nonnormal bivariate distribution with applications*, Mathematical and Computer Modelling **32**, 53–67, 2000.