SORT 28 (2) July-December 2004, 125-160

# Robust estimation and forecasting for beta-mixed hierarchical models of grouped binary data

M. A. Pashkevich, Yu. S. Kharin*

*Belarusian State University*

**Abstract**

The paper focuses on robust estimation and forecasting techniques for grouped binary data with misclassified responses. It is assumed that the data are described by the beta-mixed hierarchical model (the beta-binomial or the beta-logistic), while the misclassifications are caused by the stochastic additive distortions of binary observations. For these models, the effect of ignoring the misclassifications is evaluated and expressions for the biases of the method-of-moments estimators and maximum likelihood estimators, as well as expressions for the increase in the mean square error of forecasting for the Bayes predictor are given. To compensate the misclassification effects, new consistent estimators and a new Bayes predictor, which take into account the distortion model, are constructed. The robustness of the developed techniques is demonstrated via computer simulations and a real-life case study.

*MSC:* 62F35

*Keywords:* grouped binary data, distortions, hierarchical models, beta-binomial, beta-logistic, robust, estimation, forecasting

## 1 Introduction

Grouped binary data frequently arise in longitudinal studies that are carried out over a group of similar objects (Diggle *et al.* 2002). A natural way to describe this kind of data is using the binomial model (Collet 2002). However, the binomial model often leads to inaccurate statistical inference due to the so called "over-dispersion" effects (Brooks 2001). These effects may occur for two main reasons (Neuhaus 2002): (i) intergroup correlation, i.e. violation of the independence assumption of the experiment outcomes

* *Address for correspondence:* Department of Mathematical Modeling and Data Analysis. Faculty of Applied Mathematics and Computer Science. Belarusian State University. 4 F.Skaryny Ave., Minsk 220050, Belarus. pma@omegasoftware.com, kharin@bsu.by

Received: October 2003
Accepted: June 2004

for a particular object, and (ii) intragroup correlation caused by the heterogeneity among objects. So, special “random effects” models are used to describe the heterogeneity and correlated outcomes (Coull and Agresti 2000).

The beta-mixed hierarchical models of grouped binary data are widely used in practical applications when information about experiment conditions is not available. The most popular models of this class are the beta-binomial model (BBM) that supposes that the data on object properties are not available, and the beta-logistic model (BLM) that supposes that they are known. The BBM was originally proposed by Pearson (1925), formalized by Skellam (1948) and is associated with many useful results in applied statistics due to its conjugate property (Prentice 1988) that allows avoiding numerical integration while using Bayes approach for forecasting of response probabilities (Slaton *et al.* 2000). The BLM is an extension of the BBM that was proposed by Heckman and Willis (1977); it is widely used in economics, biometrics, political sciences and other applications (Pfeifer 1998; Nathan 1999).

In real life, the observed binary outcomes are often misclassified (Neuhaus 1999), and the classical statistical procedures that are optimal for the hypothetical model may lose their “good” properties under distortions (Kharin 1996). Hence, it is important to analyze the sensitivity of the classical estimators and predictors w.r.t. response misclassifications and, if needed, to develop new statistical procedures that are robust to these distortions (Huber 1981; Hampel *et al.* 1986). Although a number of papers have been published on robustness of the linear mixed model (Gill 2001), logistic regression (Kordzakhia *et al.* 2001), binomial model (Ruckstuhl and Welsh 2001), inference for dichotomous survey data (Gaba and Winkler 1992), and on the Bayesian identifiability problem of multinimial data with misclassifications (Swarzt *et al.* 2004), these results can not be directly applied to the grouped binary data due to their specific property.

The literature review shows that little research has been done on investigation the robustness issue for the special models of the grouped binary data. The major contribution to this domain has been done by Neuhaus, who has extended his general results for the binary regression models under response misclassifications (Neuhaus 1999) to the clustered and longitudinal binary data case. In his recent work, Neuhaus (2002) obtained expressions for the parameter bias and developed methods for consistent estimation for the population-averaged models (Liang and Zeger 1986). He also examined a special case of the cluster-specific models (Zeger and Karim 1991), the logistic normal model, which is an extension of the logistic regression to the grouped binary data case. However, as noted by Neuhaus (2002), “the derivation of bias expressions for nonlogistic links will require a different approach than for the logistic” since the specific property of the logistic link function was used to obtain the expressions.

This paper focuses on the robustness issues for the beta-mixed hierarchical models under stochastic additive distortions of binary observations. These models belong to the cluster-specific type but have not been addressed in the related works yet. The remainder of the paper is organized as follows. Section 2 is devoted to the problem statement and definition of the related mathematical models. Section 3 concentrates on the robust estimation of the beta-binomial model parameters, while Section 4 deals with the same problem for the beta-logistic model. Section 5 is dedicated to the robust forecasting based on the beta-mixed hierarchical models (both beta-binomial and beta-logistic ones). Section 6 presents an application example and evaluation of the developed methods for a real-life case study. Finally, Section 7 summarizes the main contributions of the paper.

## 2 Mathematical models and research problems

Let us consider $k$ clusters with the covariates $Z_i \in R^m, i = 1, \ldots, k$, and let $B_i = (B_{i1}, B_{i2}, \ldots, B_{in_i}) \in \{0, 1\}^{n_i}$ be the binary responses of $n_i$ Bernoulli trials over the cluster $i$. Let us also assume that the following two assumptions hold.

$A1$. Within the cluster $i$, the success probability $p_i$ is a random variable that follows the beta distribution with the true unknown parameters $\alpha_i^0 = f_\alpha(Z_i)$, $\beta_i^0 = f_\beta(Z_i)$, where $f_\alpha(.) : R^m \to R^+$, $f_\beta(.) : R^m \to R^+$.

$A2$. Random variables $p_1, p_2, \ldots, p_k$ are independent in total.

Let us refer to the defined above data model as the beta-mixed hierarchical model of the grouped binary data. In this paper, we focus on two models of this type that are frequently used in practical applications (the beta-binomial and the beta-logistic), which are specified as follows:

$$\text{BBM:} \quad f_\alpha(Z_i) = \alpha^0, \quad f_\beta(Z_i) = \beta^0, \quad n_i = n; \\ \text{model parameters: } n \in N, \quad \alpha^0, \beta^0 \in R \,.$$

$$\text{BLM:} \quad f_\alpha(Z_i) = \exp(Z_i^T a^0), \;\; f_\beta(Z_i) = \exp(Z_i^T b^0); \\ \text{model parameters: } n_1, \ldots, n_k \in N, \quad a^0, b^0 \in R^m.$$

For the BBM, it is assumed that the number of Bernoulli trials $n_i = n$ is the same for all clusters and $n$ is known a priori. Estimation of the remaining BBM parameters $\alpha^0, \beta^0$ is performed (Tripathi *et al.* 1994) using the method of moments (explicit expressions) or the method of maximum likelihood (numerical algorithm). For the BLM, the number of Bernoulli trials $n_i$ may vary across the clusters and is also known a priori, while the other parameters $a^0, b^0$ are estimated using the maximum likelihood numerical algorithm (Slaton *et al.* 2000).

One of the main problems for the grouped binary data that is strongly motivated by practical applications, is the forecasting of the success probabilities $p_1, \ldots, p_k$ for for the future trials using the past binary outcomes $B = \{B_1, \ldots, B_k\}$ obtained for small sample sizes $n_i$ that are too small to have accurate traditional estimator $\hat{p}_i = n_i^{-1} x_i^0$ (Collet 2002). For the beta-mixed hierarchical models, this problem is solved via the Bayes predictor function (Diggle *et al.* 2002)

$$\tilde{p}_i(x_i^0) = (\alpha_i^0 + x_i^0)/(\alpha_i^0 + \beta_i^0 + n_i), \tag{1}$$

where $x_i^0 = \sum_{j=1}^{n_i} B_{ij}, \quad i = 1, 2, \ldots, k$, are the sums of the binary outcomes within the cluster. This predictor ensures the minimal mean square error of forecasting when the consistent estimators of the model parameters $\{\alpha_i^0, \beta_i^0\}$ are used.

Suppose now that the original binary data $B$ are contaminated by the stochastic additive binary distortions $\{\eta_{ij}\}$, and we observe the distorted binary responses $\tilde{B}$

$$\tilde{B}_{ij} = B_{ij} \oplus \eta_{ij} \tag{2}$$

with the misclassifications defined as

$$\mathbf{P}\{\tilde{B}_{ij} = 1 | B_{ij} = 0\} = \varepsilon_0, \quad \mathbf{P}\{\tilde{B}_{ij} = 0 | B_{ij} = 1\} = \varepsilon_1, \tag{3}$$

where $\oplus$ is the modulo 2 sum, and $\varepsilon_0, \varepsilon_1 \ll 1$ are the distortion levels which can be either known or unknown (Copas 1988). In this settings, two main research problems arise:

(i) Evaluation of the effects of ignoring the misclassifications for the classical model parameter estimation techniques and response probability forecasting methods.

(ii) Construction of new estimation and prediction methods, which take into account the distortion model and compensate the misclassification effect.

In the remaining sections, these problems are solved separately for the BBM and BLM parameter estimation, while the forecasting is examined and enhanced simultaneously for both of them. For the first problem, the estimation bias and the increase in the mean square error of forecasting are evaluated via asymptotic expansions. For the second one, new estimation and forecasting methods, which are based on the obtained probability distribution of the distorted data, are proposed.

It should be noted that for the BBM (Section 3), the paper considers the case of equal group sizes since it is typical for many application areas that exploit this model. The assumption $n_i = n$ allows obtaining simple expressions and helps to develop intuition about the distortions influence on the BBM inference. However, the results for the BBM with different $\{n_i\}$ can be easily obtained as a special case of the BLM results (Section 4), where the covariates $\{Z_i\}$ are the same for all clusters.

For further convenience, let us introduce the following notation: MM-estimator – the method of moments estimator, ML-estimator– the method of maximum likelihood estimator, $o(\varepsilon)$, $O(\varepsilon)$ –Landau symbols for $\varepsilon \to 0$, $Y_n = O_P(Z_n)$– probability Landau symbol for random sequences $Y_n, Z_n \in R$. The detailed definition of $O_P(.)$ and the proofs of theorems are given in Mathematical Appendix.

## 3 Robust estimation of the beta-binomial model

**Distorted beta-binomial distribution.** Let $x_i$ be the number of successes for the $i$-th cluster: $x_i = \sum_{j=1}^{n_i} \tilde{B}_{ij}, \quad i = 1, \ldots, k$. The following theorem defines the probability distribution of the random variable $x_i$ under the distortions (2), (3).

**Theorem 1** *The probability distribution of the distorted beta-binomial random variable $x_i$ can be represented as a weighted sum*

$$P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1) = \sum_{s=0}^{n} w_{rs}(\varepsilon_0, \varepsilon_1) \cdot P_s^0(\alpha, \beta), \tag{4}$$

*where $\{P_s^0\}$ are the non-distorted probabilities for the BBM with the parameters $n, \alpha, \beta$*

$$P_s^0(\alpha, \beta) = \binom{n}{s} \frac{B(\alpha + s, \beta + n - s)}{B(\alpha, \beta)},$$

*$B(.)$ is the complete beta function, and the weights for the distortion levels $\varepsilon_0, \varepsilon_1$ are computed as*

$$w_{rs}(\varepsilon_0, \varepsilon_1) = \sum_{l=\max(s,r)}^{\min(n,s+r)} \binom{s}{l-r} \binom{n-s}{l-s} \varepsilon_0^{l-s} (1-\varepsilon_0)^{n-l} \varepsilon_1^{l-r} (1-\varepsilon_1)^{s+r-l}, \quad s, r = 0, 1, \ldots, n.$$

Using this theorem, it can be proved that the mean and variance of the distribution (4) are

$$\mathbf{E}\{x_i\} = \varepsilon_0 \frac{n\beta}{\alpha + \beta} + (1-\varepsilon_1) \frac{n\alpha}{\alpha + \beta}, \qquad \mathbf{V}\{x_i\} = \varepsilon_0(1-\varepsilon_0) \frac{n\beta}{\alpha + \beta} + \varepsilon_1(1-\varepsilon_1) \frac{n\alpha}{\alpha + \beta} + (1-\varepsilon_0-\varepsilon_1)^2 \cdot V_0,$$

where $V_0 = (n\alpha\beta(\alpha+\beta+n))/((\alpha+\beta)^2(\alpha+\beta+1))$ is the variance of the non-distorted BBM. Let us refer to the distribution (4) as the distorted beta-binomial distribution (DBBD) with the parameters $n$, $\alpha$, $\beta$, $\varepsilon_0$, $\varepsilon_1$.

As follows from the theorem proof (see Appendix), the weights $w_{rs}$ can be treated as the probabilities that the distorted value $r$ was originated from the non-distorted sum of the binary outcomes $s$. It should be noted that when $\varepsilon_0 = \varepsilon_1 = 0$, the proposed distribution (4) is identical to the classical beta-binomial distribution (BBD) with the parameters $n$, $\alpha$, $\beta$, and the weight matrix $W = (w_{rs})$ is the identity one. If the distortion levels are small, the matrix $W$ can be approximated by the asymptotic expansion

$$W(\varepsilon_0, \varepsilon_1) = I + W'_{\varepsilon_0} \cdot \varepsilon_0 + W'_{\varepsilon_1} \cdot \varepsilon_1 + o(\varepsilon_0, \varepsilon_1), \tag{5}$$

where $I$ is the identity matrix, and the matrices $W'_{\varepsilon_0}$, $W'_{\varepsilon_1}$ are calculated as

$$W'_{\varepsilon_0} = \begin{pmatrix} -n & 0 & 0 & \dots & 0 \\ n & -(n-1) & 0 & \dots & 0 \\ 0 & n-1 & -(n-2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \qquad W'_{\varepsilon_1} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ 0 & 0 & -2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -n \end{pmatrix}.$$

The expression (5) allows obtaining the following asymptotic relation between the distorted $P_r(\varepsilon_0, \varepsilon_1)$ and the original $P_r^0$ probabilities

$$P_r(\varepsilon_0, \varepsilon_1) = P_r^0 + \Big((n-r+1)P_{r-1}^0 - (n-r)P_r^0\Big) \cdot \varepsilon_0 + \Big((r+1)P_{r+1}^0 - rP_r^0\Big) \cdot \varepsilon_1 + o(\varepsilon_0, \varepsilon_1), \tag{6}$$

where $P_{-1}^0 = P_{n+1}^0 = 0$. This expression can be employed to assess the sensitivity of the beta-binomial distribution to the distortions (2), (3). In the following subsection, the result of Theorem 1 and the expression (6) are used to evaluate the sensitivity of the classical BBM estimators.

**Robustness of the classical estimators.** Let $\alpha^0, \beta^0$ be the true unknown values of the BBM parameters, and let $\Delta\tilde{\alpha}(\varepsilon_0, \varepsilon_1)$, $\Delta\tilde{\beta}(\varepsilon_0, \varepsilon_1)$ be the biases of the parameter estimators that ignore the misclassifications with the levels $\varepsilon_0, \varepsilon_1$. The following theorems evaluate the robustness of the classical MM and ML-estimators via their biases w.r.t. the distortion levels.

**Theorem 2** *The bias of the classical MM-estimator of the BBM parameters, which ignores the misclassifications, satisfies the following asymptotic expansion*

$$\begin{pmatrix} \Delta\widetilde{\alpha}_{MM} \\ \Delta\widetilde{\beta}_{MM} \end{pmatrix} = \begin{pmatrix} \alpha^0 + 2\beta^0 + 1 & \alpha^0(\alpha^0+1)/\beta^0 \\ \beta^0(\beta^0+1)/\alpha^0 & 2\alpha^0 + \beta^0 + 1 \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} + \begin{pmatrix} o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \\ o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \end{pmatrix}. \tag{7}$$

**Theorem 3** *The bias of the classical ML-estimator of the BBM parameters, which ignores the misclassifications, satisfies the asymptotic expansion*

$$\begin{pmatrix} \Delta\widetilde{\alpha}_{ML} \\ \Delta\widetilde{\beta}_{ML} \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}^{-1} \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} + \begin{pmatrix} o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \\ o(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k}) \end{pmatrix}, \tag{8}$$

*where explicit expressions for the matrices H, G are given in Mathematical Appendix.*

As follows from these theorems, the classical MM and ML-estimators of the BBM parameters become biased and inconsistent under the distortions. Expressions (7), (8) allow assessing the sensitivity of these estimators to the misclassifications (3). Let us now construct consistent and unbiased estimators that take into account the distortion model (2), (3).

**Robust estimation in the case of known distortion levels.** Let us consider the case when the distortion levels $\varepsilon_0, \varepsilon_1$ are known a priori. Denote the empirical moments of the order $r$ as $m_r^* = k^{-1} \sum_{i=1}^{k} x_i^r$. The following theorems define consistent and asymptotically unbiased MM and ML-estimators for the case of known $\varepsilon_0, \varepsilon_1$.

**Theorem 4** *The consistent and asymptotically unbiased MM-estimator, which takes into account the distortion model (2), (3), is expressed as*

$$\hat{\alpha}_{MM} = \frac{\delta_\alpha(m_1^*, \varepsilon_0) \cdot \mu(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}{\Delta(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}, \qquad \hat{\beta}_{MM} = \frac{\delta_\beta(m_1^*, \varepsilon_1) \cdot \mu(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}{\Delta(m_1^*, m_2^*, \varepsilon_0, \varepsilon_1)}, \tag{9}$$

*where*

$$\delta_\alpha = m_1^* - n\varepsilon_0, \quad \delta_\beta = n - m_1^* - n\varepsilon_1, \quad \mu = m_1^* n - m_2^* - (\varepsilon_0 \delta_\beta + m_1^* \varepsilon_1)(n-1),$$

$$\Delta = (1-\varepsilon_1-\varepsilon_0)\left(m_2^* n - m_1^* n - m_1^{*2}(n-1)\right).$$

**Theorem 5** *The consistent and asymptotically unbiased ML-estimator, which takes into account the distortion model (2), (3), can be derived by applying the classical ML-estimator to the filtered empirical probabilities*

$$\hat{P}_r^0 = \sum_{l=0}^{n} \upsilon_{rl}(\varepsilon_0, \varepsilon_1) \cdot \hat{P}_l(\varepsilon_0, \varepsilon_1), \tag{10}$$

*where* $\{\hat{P}_0, \ldots, \hat{P}_n\}$ *is the empirical probability distribution of the distorted sample* $\{x_1, x_2, \ldots, x_k\}$*, and* $\upsilon_{rl}$ *are the elements of the inverted weight matrix W from Theorem 1:* $V = (\upsilon_{rl}) = W^{-1}, \;\; det(W) \neq 0$.

Let us refer to the above estimators as the modified MM-estimator (MMM-estimator) and the modified ML-estimator (MML-estimator) respectively. It should be noted that the filtration approach (Theorem 5) is not limited to the maximum likelihood technique, it can also be used together with other known estimation methods developed for the classical (non-distorted) beta-binomial distribution. A good review of these methods can be found in (Tripathi *et al.*, 1994).

**Robust estimation in the case of unknown distortion levels.** Let us now consider a general case when both the BBM parameters $\alpha, \beta$ and the distortion levels $\varepsilon_0$, $\varepsilon_1$ are unknown. For simultaneous consistent estimation of $\alpha, \beta$ and $\varepsilon_0, \varepsilon_1$, two numerical algorithms are proposed; the first employs the method of moments and the second utilizes the maximum likelihood approach.

For the method of moments, the simultaneous estimation problem can be reduced to the solution of the following system of two nonlinear equations for the third and fourth order moments

$$m_3^* = m_3(\alpha(\varepsilon_0,\varepsilon_1),\beta(\varepsilon_0,\varepsilon_1),\varepsilon_0,\varepsilon_1), \quad m_4^* = m_4(\alpha(\varepsilon_0,\varepsilon_1),\beta(\varepsilon_0,\varepsilon_1),\varepsilon_0,\varepsilon_1), \tag{11}$$

where the functions $\alpha(\varepsilon_0,\varepsilon_1)$, $\beta(\varepsilon_0,\varepsilon_1)$ are expressed explicitly (see Theorem 4) from the equations for the first and second order moments

$$m_1^* = m_1(\alpha,\beta,\varepsilon_0,\varepsilon_1), \quad m_2^* = m_2(\alpha,\beta,\varepsilon_0,\varepsilon_1). \tag{12}$$

Here $m_r^* = k^{-1}\sum_{i=0}^{k} x_i^r$, $r = 1,2,3,4$; while $m_r(\alpha,\beta,\varepsilon_0,\varepsilon_1)$ are the corresponding theoretical moments for the DBBD with the parameters $n,\alpha,\beta,\varepsilon_0,\varepsilon_1$ that can be computed using Theorem 1. To solve the equations (11), let us apply the modified Newton method. Denote by $J_0^c$ the 2×2 Jacobi matrix of the system (11) on the condition that the equations for the first two moments (12) hold. Then the iterative procedure for the solution of (11) is expressed as

$$\begin{pmatrix} \varepsilon_0^{t+1} \\ \varepsilon_1^{t+1} \end{pmatrix} = \begin{pmatrix} \varepsilon_0^{t} \\ \varepsilon_1^{t} \end{pmatrix} + \lambda \cdot (J_0^c)^{-1} \begin{pmatrix} m_3^* - m_3(\alpha(\varepsilon_0^t,\varepsilon_1^t),\beta(\varepsilon_0^t,\varepsilon_1^t),\varepsilon_0^t,\varepsilon_1^t) \\ m_4^* - m_4(\alpha(\varepsilon_0^t,\varepsilon_1^t),\beta(\varepsilon_0^t,\varepsilon_1^t),\varepsilon_0^t,\varepsilon_1^t) \end{pmatrix}, \tag{13}$$

where $\lambda \in (0,1]$ is the algorithm parameter that ensures the convergence for large distortion levels $\varepsilon_0,\varepsilon_1$ (Demidovich and Maron 1970). All expressions required for the numerical implementation of the procedure (13) are given in the Mathematical Appendix. As follows from the numerical experiments, the usual value $\lambda = 1$ (typical for the classical Newton technique) provides poor convergence, so it is prudent to start iterations with rather low $\lambda$ and gradually increase it so that it becomes close to 1 in the neighborhood of the desired solution. It can be done using the recursive sequence $\lambda_{t+1} = \lambda_t \cdot (1-\theta) + \theta$, where $\lambda_0$ and $\theta$ are the tuning parameters. During the computer simulations that will be discussed below, the authors used the following values: $\lambda_0 = 0.1$, $\theta = 0.05$. Let us refer to the estimates of the model parameters $\alpha,\beta$ and the distortion levels $\varepsilon_0,\varepsilon_1$ obtained using the procedure (13) as the MMS-estimates.

For the maximum likelihood approach, the simultaneous estimation is reduced to the following constrained maximization problem

$$l(\alpha,\beta,\varepsilon_0,\varepsilon_1) = \sum_{r=0}^{n} f_r \ln(P_r(\alpha,\beta,\varepsilon_0,\varepsilon_1)) \to \max_{\alpha,\beta,\varepsilon_0,\varepsilon_1}, \quad \alpha,\beta \in R^+, \quad \varepsilon_0,\varepsilon_1 \in [0,1], \tag{14}$$

where $\{f_0, f_1, \ldots, f_n\}$ are the frequencies for the distorted sample $\{x_1, x_2, \ldots, x_k\}$, and the explicit expressions for the distorted beta-binomial probabilities $P_r(.)$ are given in

Theorem 1. This maximization problem is solved using the modification of the steepest descent method. All the expressions required for the numerical implementation are given in Mathematical Appendix. Let us refer to the estimates of $\alpha, \beta$ and $\varepsilon_0, \varepsilon_1$ obtained from (14) as the MLS-estimates.

**Computer simulations.** To demonstrate the robustness of the proposed estimators of the BBM parameters, a series of four computer simulations was done. It was assumed that the true values of the model parameters were $\alpha^0 = 0.5$, $\beta^0 = 9.5$, $n = 10$. These values are typical for the application area that the authors deal with (see Section 6).

*Experiment 1*. This experiment was dedicated to assessing the sensitivity of the beta-binomial distribution to the distortions (Theorem 1). There were generated $k = 1000$ realizations of the random variable from the DBBD with the parameters $n, \alpha^0, \beta^0$ and the distortion levels $\varepsilon_0 = 0.01$, $\varepsilon_1 = 0.02$. For the generated sample, there were computed the empirical probabilities $P_r^*$, $r = 0, 1, \ldots, n$, as well as the sample mean and variance. Also, there were calculated the weight matrix $W$, the theoretical probabilities $P_r$ and $P_r^0$, the approximate values $P_r^a$ for $P_r$ (the asymptotic expansion (6)), and the theoretical mean and variance for the BBD and DBBD.

As follows from the experiment results (Tables 1-3), the original beta-binomial distribution is quite sensitive to the distortions. For example, the relative difference between the non-distorted $P_r^0$ and distorted $P_r$ probabilities can go up to 24.9%, and the mathematical expectation and variance can differ by 17.0% and 3% respectively. The corresponding weight matrix $W$ (see Table 3) has the dominated leading diagonal and the adjacent elements, that explains why the linearized expressions (6) provide an accurate enough approximation of the probabilities $P_r$. This result validates using of stochastic expansions for assessing the sensitivity of the classical estimation and prediction techniques with respect to the distortion levels.

***Table 1***: *Comparison of the original, distorted and empirical mean and variance.*

| Distribution type | Mean | Variance |
|---|---|---|
| Classical beta-binomial distribution | 0.500 | 0.929 |
| Distorted beta-binomial distribution | 0.585 | 0.957 |
| Empirical distribution | 0.577 | 0.943 |

***Table 2***: *Comparison of the original, distorted and empirical probabilities for the BBM.*

| $r$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\times 10^{-1}$ | $\times 10^{-1}$ | $\times 10^{-2}$ | $\times 10^{-2}$ | $\times 10^{-2}$ | $\times 10^{-3}$ | $\times 10^{-3}$ | $\times 10^{-4}$ | $\times 10^{-4}$ | $\times 10^{-5}$ | $\times 10^{-6}$ |
| $P_r^0$ | 6.93 | 1.87 | 7.23 | 2.92 | 1.15 | 4.30 | 1.46 | 4.34 | 1.06 | 1.91 | 1.91 |
| $P_r$ | 6.30 | 2.34 | 8.40 | 3.24 | 1.25 | 4.54 | 1.50 | 4.36 | 1.04 | 1.80 | 1.73 |
| $P_r^a$ | 6.28 | 2.39 | 8.22 | 3.21 | 1.24 | 4.52 | 1.50 | 4.35 | 1.03 | 1.80 | 1.72 |
| $P_r^*$ | 6.32 | 2.34 | 8.33 | 3.19 | 1.17 | 5.00 | 0.95 | 2.50 | 1.50 | 0.00 | 1.91 |

***Table 3***: *Elements of the weight matrix W for the distortion levels* $\varepsilon_0 = 0.01$*,* $\varepsilon_1 = 0.02$*.*

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.9044 | 0.0183 | 0.0004 | $\sim 10^{-5}$ | $\sim 10^{-7}$ | $\sim 10^{-9}$ | $\sim 10^{-10}$ | $\sim 10^{-12}$ | $\sim 10^{-14}$ | $\sim 10^{-15}$ | $\sim 10^{-17}$ |
| 1 | 0.0914 | 0.8969 | 0.0362 | 0.0011 | $\sim 10^{-5}$ | $\sim 10^{-6}$ | $\sim 10^{-8}$ | $\sim 10^{-10}$ | $\sim 10^{-11}$ | $\sim 10^{-13}$ | $\sim 10^{-14}$ |
| 2 | 0.0042 | 0.0815 | 0.8891 | 0.0538 | 0.0022 | 0.0001 | $\sim 10^{-6}$ | $\sim 10^{-7}$ | $\sim 10^{-9}$ | $\sim 10^{-11}$ | $\sim 10^{-12}$ |
| 3 | 0.0001 | 0.0033 | 0.0717 | 0.8811 | 0.0710 | 0.0036 | 0.0001 | $\sim 10^{-5}$ | $\sim 10^{-7}$ | $\sim 10^{-8}$ | $\sim 10^{-10}$ |
| 4 | $\sim 10^{-6}$ | 0.0001 | 0.0025 | 0.0621 | 0.8727 | 0.0879 | 0.0053 | 0.0003 | $\sim 10^{-5}$ | $\sim 10^{-7}$ | $\sim 10^{-8}$ |
| 5 | $\sim 10^{-8}$ | $\sim 10^{-6}$ | 0.0001 | 0.0019 | 0.0527 | 0.8641 | 0.1044 | 0.0074 | 0.0004 | $\sim 10^{-5}$ | $\sim 10^{-6}$ |
| 6 | $\sim 10^{-10}$ | $\sim 10^{-8}$ | $\sim 10^{-6}$ | $\sim 10^{-5}$ | 0.0013 | 0.0435 | 0.8551 | 0.1206 | 0.0097 | 0.0006 | $\sim 10^{-5}$ |
| 7 | $\sim 10^{-12}$ | $\sim 10^{-10}$ | $\sim 10^{-8}$ | $\sim 10^{-7}$ | $\sim 10^{-5}$ | 0.0009 | 0.0344 | 0.8460 | 0.1363 | 0.0124 | 0.0008 |
| 8 | $\sim 10^{-15}$ | $\sim 10^{-13}$ | $\sim 10^{-11}$ | $\sim 10^{-9}$ | $\sim 10^{-7}$ | $\sim 10^{-5}$ | 0.0005 | 0.0256 | 0.8366 | 0.1517 | 0.0153 |
| 9 | $\sim 10^{-17}$ | $\sim 10^{-15}$ | $\sim 10^{-13}$ | $\sim 10^{-11}$ | $\sim 10^{-9}$ | $\sim 10^{-8}$ | $\sim 10^{-6}$ | 0.0003 | 0.0169 | 0.8269 | 0.1667 |
| 10 | $\sim 10^{-20}$ | $\sim 10^{-18}$ | $\sim 10^{-16}$ | $\sim 10^{-14}$ | $\sim 10^{-12}$ | $\sim 10^{-10}$ | $\sim 10^{-8}$ | $\sim 10^{-6}$ | 0.0001 | 0.0083 | 0.8171 |

*Experiment 2*. This experiment was devoted to assessing the bias of the classical BBM parameter estimators that ignore the misclassifications (Theorems 2, 3). There were generated 100 independent random samples of size $k = 1000$ from the BBM with the parameters $n, \alpha^0, \beta^0$. It was assumed that $\varepsilon_0 = \varepsilon_1 \in [\,0;\ 0.02\,]$ and they varied with the step 0.002, and each sample was contaminated according to the distortion model (2), (3). For each distorted sample and for each value of the distortion level, the classical MM and ML methods were applied. Then, for all values of $\varepsilon_0, \varepsilon_1$, the 95%-confidence intervals of the $\alpha, \beta$ estimates were computed (using the common technique, which assumes that the estimates follow the normal distribution). Finally, for the same distortion levels, the theoretical biases were obtained using the stochastic expansions (7), (8).

The results of the experiment are presented in Figure 1, where $\rho(.)$ is the relative bias (i.e. $\Delta\alpha/\alpha^0$ or $\Delta\beta/\beta^0$). As follows from the figure, the stochastic expansions (7), (8) provide good approximation of the parameters biases caused by the distortions with the levels $\varepsilon_0, \varepsilon_1 \leq 0.01$. Besides, the classical estimators are quite sensitive to the distortions. For example, for the distortion levels $\varepsilon_0 = \varepsilon_1 = 0.01$, the relative errors for the parameters $\alpha, \beta$ are respectively 50%, 22.7% for the the MM-estimator and 52.7%, 28.0% for the ML-estimator.

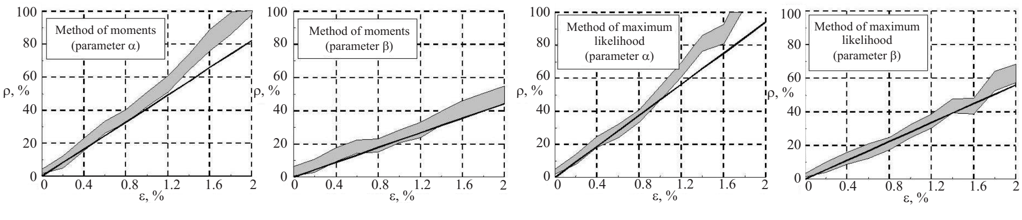


***Figure 1***: *The biases of the classical MM- and ML-estimators of the BBM parameters, which ignore the misclassifications: gray tubes –experimental 95% confidence intervals; solid lines– approximation via the asymptotic expansions (7), (8);* $\rho$ *–the relative bias,* $\varepsilon$*– the distortion level (*$\varepsilon_0$*=*$\varepsilon_1$*).*

However, for practical applications, it is also important to analyze the sensitivity of another BBM parametrization (Prentice 1986): $\pi = \alpha/(\alpha+\beta)$, $\gamma = 1/(\alpha+\beta)$, where $\pi$ is the average response probability, and $\gamma$ is a measure of the inter-group correlation. For this parametrization, the relative errors of the MM and ML-estimators for the parameters $\pi, \gamma$ are 20.9%, 19.4% and 17.7%, 22.6% respectively. It means that ignoring response misclassifications leads to quite large errors when assessing both the average response probability for the clusters and the inter-group correlation between units. This numerical result emphasizes the importance of the research topic and motivates development of new robust estimators, which take the distortion model into account.

It should be noted that Neuhaus (1999, 2002) performed similar computer simulations for the binary regression, as well as for the population-averaged and the mixed-effects logistic models. In his simulation, Neuhaus was interested in the bias of the regression coefficients and made a conclusion that the biases due to response misclassifications were negligible for small values of the distortion levels and were substantial only when $\varepsilon_0, \varepsilon_1 \geq 0.10$. Since our experiments yielded qualitatively different results (see Figure 1), this fact should be explained in details.

For the comparison purposes, the beta-mixed hierarchical model considered in this paper (both BBM and BLM) can be reformulated as a special case of the generalized linear mixed model (GLMM), which is an extension of the generalized linear model (GLM) to the longitudinal or clustered data case. The reformulation can be done by introducing dummy constant covariates for each cluster/unit, and choosing an appropriate link function and a random effects distribution. Then the regression coefficients can be considered as the beta-mixed hierarchical model parameters, and their sensitivity to the distortions can be investigated using technique employed in this paper. Hence, the above model conversion can be treated as a specific nonlinear re-parametrization of the beta-mixed hierarchical model, which leads to completely different meaning of the model parameters.

For this re-parametrization, the parameter estimator sensitivity w.r.t. the misclassifications may increase, depending on the true values of the parameter. For instance, for small values of $\pi$ (which are typical for our application area), the misclassifications essentially influence the estimate $\hat{\pi}$, since $E\{x_i/n\} = \varepsilon_0(1–\pi)+(1–\varepsilon_1)\pi$. Thus, when $\pi = 0.05$ and $\varepsilon_0 = \varepsilon_1 = 0.01$ the expectation of $x_i/n$ is equal to 0.059, i.e. misclassifications cause 18% increase of the corresponding parameter value. This justifies the qualitative difference of the Neuhaus' and ours simulation results.

Therefore, the obtained results show that the beta-binomial model parameter estimators are less robust to the response misclassifications compared to the estimators for the models investigated by Neuhaus. This emphasizes the research topic importance and motivates development of robust estimators for the BBM. It should be also noted that the robust estimation approach for the logistic-normal model that was employed by Neuhaus (2002) can not be applied to the BBM since he used specific properties of the logistic link function that the beta-binomial distribution does not possess.

*Experiment 3*. This experiment was aimed at the performance evaluation of the proposed robust estimators in the case of known distortion levels (Theorems 4, 5). It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.01$, and the developed MMM and MML-estimators were compared to the classical MM and ML-estimators by assessing the biases, standard deviations, and histograms. As follows from the experiment results (Figure 2), the proposed estimation methods allow essentially decreasing the bias of the $\alpha, \beta$ estimates and lead to the smaller standard deviation while compared to the classical estimators. In particular, the MMM-estimator yields the relative biases 2.0%, 2.1% for the parameters $\alpha, \beta$ respectively against 47.7%, 25.2% obtained by applying the classical MM technique. The MML-estimator ensures the relative biases 0.9%, 1.1% in contrast to 54.2%, 30.3% for the classical ML method. These results confirm the robust performance of the proposed estimators.

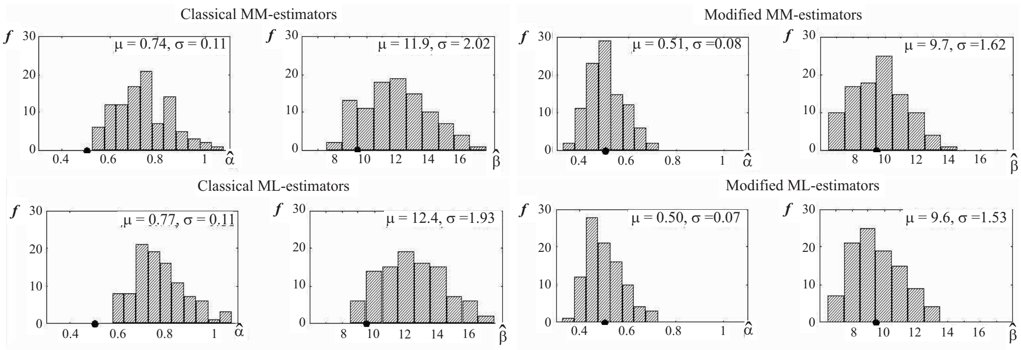


***Figure 2***: *Histograms of the classical and proposed estimators of the BBM parameters for known distortion levels: $f$ –empirical frequency, $\mu$– sample mean, $\sigma$ –sample standard deviation; the circles denote the true parameter values.*

*Experiment 4*. This experiment focused on the performance evaluation of the proposed robust estimators in the case of unknown distortion levels. It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.01$, and the developed MMS and MLS-estimators were compared to the classical MM and ML-estimators by assessing the biases and standard deviations. As follows from the experiment results (Table ), the proposed estimation techniques allow essentially decreasing the bias of the $\alpha, \beta$ estimates, while the standard deviation increases compared to the classical estimators. In particular, the MMS-estimator yields the relative biases 2.0%, 3.2% for the parameters $\alpha, \beta$ respectively against 46.0%, 24.3% obtained by applying the classical MM technique. The MLS-estimator ensures the relative biases 6.0%, 1.2% in contrast to 52.0%, 29.3% for the classical ML method. On the other hand, the standard deviation increases up to twice compared to the classical methods that ignore the misclassifications. This effect is caused by the identification of two extra parameters $\varepsilon_0, \varepsilon_1$ in addition to $\alpha, \beta$ that normally leads to extra variation.

Advantages of the developed methods were also confirmed by additional numerical research aimed at the identifiability analysis, which was based on computing of the determinant and condition number for the relevant Jacobi matrices. For the MMS-estimator, there were examined both the full 4×4 Jacobian of the system (11), (12) and the reduced 2×2 Jacobian, which is used in the numerical procedure (13). During the simulation, the determinant of the full Jacobian was far from zero and varied from 0.01 to 0.10 that confirms the identifiability. However, the corresponding condition number was rather high (from $5.8 \cdot 10^5$ to $1.4 \cdot 10^6$), that validates using of the proposed iterative procedure (13), which employs inversion of the 2×2 matrix with much better condition number (from 55.4 to 73.8). For the MLS-estimator, there was examined the 4×4 matrix of the second derivatives for the log-likelihood function (14). Its determinant was greater then $10^5$ that indicates the identifiability of all model parameters. But the corresponding condition number varied from $5.2 \cdot 10^5$ to $7.8 \cdot 10^8$ that explains slow convergence of the optimization routine (approximately 85 times slower then for the MMS-estimator) due to the ravine structure of the objective function. Nevertheless, the MLS technique gives better estimation results in comparison with the MMS (in 48% of simulation runs, the MLS biases were smaller then the MMS biases for all four parameters, in 27% of runs –for three parameters, in 19% of runs– for two parameters, in 5% of runs –for one parameter, and only in 1% of runs the MMS dominated over the MLS for all the parameters). These results confirm both the identifiability and the robust performance of the developed estimators.

***Table 4***: *Comparison of the classical and proposed estimators of the BBM for unknown distortion levels.*

| Parameter | $\alpha$ (true value 0.5) | | | | $\beta$ (true value 9.5) | | | |
|---|---|---|---|---|---|---|---|---|
| Method | MM | ML | MMS | MLS | MM | ML | MMS | MLS |
| Mean | 0.73 | 0.76 | 0.49 | 0.53 | 11.81 | 12.28 | 9.20 | 9.39 |
| Standard deviation | 0.12 | 0.11 | 0.21 | 0.22 | 2.01 | 1.97 | 2.63 | 0.71 |

## 4 Robust estimation of the beta-logistic model

**Robustness of the classical ML-estimator.** Let $a^0, b^0$ be the true unknown values of the BLM parameters, and let $\Delta\tilde{a}(\varepsilon_0, \varepsilon_1)$, $\Delta\tilde{b}(\varepsilon_0, \varepsilon_1)$ be the biases of the parameter estimators that ignore the misclassifications with the levels $\varepsilon_0, \varepsilon_1$. The following theorems evaluate the robustness of the classical ML-estimator via its bias w.r.t. the distortion levels.

**Theorem 6** *The bias of the classical ML-estimator of the BLM parameters, which ignores the misclassifications, satisfies the following asymptotic expansion*

$$\begin{pmatrix} \Delta\tilde{a} \\ \Delta\tilde{b} \end{pmatrix} = -H^{-1}G \cdot \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \end{pmatrix} + \mathbf{1}_{2m}\left(o(\varepsilon_0, \varepsilon_1) + O_P\left(\frac{1}{\sqrt{k}}\right)\right), \quad (15)$$

*under the assumption that the covariates $Z_i$ belong to the countable set $\{\vartheta_1, \vartheta_2, \ldots, \vartheta_d\} \subset R^m$, $i = 1, 2, \ldots, k$, all vectors $\{\vartheta_q\}$ are equiprobable, and the clusters* *with the same covariates $\vartheta_q$ have equal group sizes $\check{n}_q$, where $\mathbf{1}_{2m}$ is a vector of ones of size $2m$, and $H$, $G$ are $(2m \times 2m)$, $(2m \times 2)$ matrices given in Mathematical Appendix.*

As follows from the theorem, the classical ML-estimator of the BLM parameters becomes biased and inconsistent under the distortions. Expression (15) allows assessing the sensitivity of this estimator to the misclassifications (3). Let us now propose consistent and unbiased estimators that take into account the distortion model (2).

**Robust estimation in the case of known distortion levels.** Consider the case when the distortion levels $\varepsilon_0, \varepsilon_1$ are known a priori. First, let us obtain a stochastic expansion for the biases that differs from (15) by taking into account an observed sample $X = \{x_1, x_2, \ldots, x_k\}$.

**Theorem 7** *For the observed sample $X$, the bias of the classical ML-estimator of the BLM parameters, which ignores the misclassifications, satisfies the following asymptotic expansion*

$$\begin{pmatrix} \Delta\tilde{a} \\ \Delta\tilde{b} \end{pmatrix} = J^{-1}(a^0, b^0, X) \cdot g_e(a^0, b^0, X, \varepsilon_0, \varepsilon_1) + \mathbf{1}_{2m}\left(o(\varepsilon_0, \varepsilon_1) + O_P\left(\frac{1}{\sqrt{k}}\right)\right), \qquad (16)$$

*where the $(m \times m)$-matrix $J(.)$ and the $m$-vector $g_e(.)$ are defined in Mathematical Appendix.*

Then, the expansion (16) allows constructing a bias compensating procedure for the classical ML-estimator

$$(\tilde{a}^{t+1}, \tilde{b}^{t+1})^T = (\tilde{a}^t, \tilde{b}^t)^T - \lambda \cdot J^{-1}(\tilde{a}^t, \tilde{b}^t, X) \cdot g_e(\tilde{a}^t, \tilde{b}^t, X, \varepsilon_0, \varepsilon_1), \qquad (17)$$

where $t$ is the iteration number and $\lambda$ is the algorithm parameter that ensures convergence. The parameter $\lambda$ is selected in the similar way as for the numerical procedure (13) from Section 3: $\lambda_{t+1} = \lambda_t \cdot (1-\theta) + \theta$. In the given below computer simulations, the authors used values $\lambda_0 = 0.1$, $\theta = 0.05$. Let us refer to the bias-corrected ML-estimator (17) as the modified ML-estimators (MML).

**Robust estimation in the case of unknown distortion levels.** Let us now consider a general case when both the BLM parameters $a, b$ and the distortions levels $\varepsilon_0$, $\varepsilon_1$ are unknown. For simultaneous consistent estimation of $a, b$ and $\varepsilon_0, \varepsilon_1$, a maximum likelihood based numerical algorithm is proposed.

Using results from Section 3, the log-likelihood function for the BLM that accommodates the distortion model (2), (3) may be expressed as

$$l_\varepsilon(a, b, X, \varepsilon_0, \varepsilon_1) = \sum_{i=1}^{k} \log\left(\sum_{j=0}^{n_i} w^i_{x_i j}(\varepsilon_0, \varepsilon_1) \cdot P^i_j(a, b)\right), \qquad (18)$$

where

$$P^i_j(a,b) = \binom{n_i}{j} \frac{B(\alpha_i(a)+j, \beta_i(b)+n_i-j)}{B(\alpha_i(a),\beta_i(b))}, \quad \alpha_i(a) = \exp(Z_i^T a), \quad \beta_i(b) = \exp(Z_i^T b),$$

$B(.)$ is the complete beta function, and $w^i_{x_i j}$ are the weights of the distorted beta-binomial distribution with the parameters $n_i, \alpha_i(a), \beta_i(b), \varepsilon_0, \varepsilon_1$ (see Theorem 1). Then, the simultaneous estimation of the BLM parameters and the distortion levels is reduced to the following constrained maximization problem

$$l_\varepsilon(a, b, X, \varepsilon_0, \varepsilon_1) \to \max_{a,b,\varepsilon_0,\varepsilon_1}, \quad a, b \in R^m, \quad \varepsilon_0, \varepsilon_1 \in [0, 1]. \tag{19}$$

The problem (19) is solved using the gradient descent method; all the required expressions are given in the Mathematical Appendix. Let us refer to the estimates of $a, b$ and $\varepsilon_0, \varepsilon_1$ obtained from (19) as the MLS-estimates.

**Computer simulations.** To demonstrate the robust performance of the developed methods for the estimation of the BLM, a number of computer simulations was done. It was assumed that the true values of the parameters were $a^0 = 1$, $b^0 = 2$, $\forall i, n_i = 10$, $k = 1000$, and the covariates $Z_i \in R$ were uniformly distributed on the segment [1.0; 1.1]. This range of the covariates corresponds to the intervals $\alpha \in [2.7; 3.0]$, $\beta \in [7.4; 9.0]$ that is typical for the application area the authors deal with (see Application Example). The simulations included three experiments.

*Experiment 1*. This experiment was devoted to assessing the bias of the classical ML-estimator of the BLM parameters that ignores the misclassifications (Theorem 6). There were generated 100 independent random samples of size $k = 1000$ from the BLM with the parameters $a^0, b^0$. It was assumed that $\varepsilon_0 = \varepsilon_1 \in [0;\ 0.05]$ and they varied with the step 0.01, and each sample was contaminated according to the distortion model (2), (3). For each distorted sample and for each value of the distortion level, the classical ML method was applied. Then, for all values of $\varepsilon_0, \varepsilon_1$, the 95%-confidence intervals of the $a, b$ estimates were computed (assuming that the estimates follow the normal distribution). Finally, for the same distortion levels, the theoretical biases were obtained using the stochastic expansion (15).

As follows from the experiment results (Figure 3), the stochastic expansion (15) provides good approximation of the parameters biases caused by the distortions with the levels $\varepsilon_0, \varepsilon_1 \le 0.05$. Besides, the classical ML-estimator is quite sensitive to the distortions. For example, for the distortion levels $\varepsilon_0 = \varepsilon_1 = 0.05$, the relative errors for the parameters $a, b$ are 39.2%, 12.1% respectively. It should be noted that the higher parameter biases in comparison with the results of Neuhaus (1999) for the binary regression are due to the specific nonlinear parametrization of the beta-mixed hierarchical models (for details, see the above discussion in the Computer Simulation subsection for the BBM).

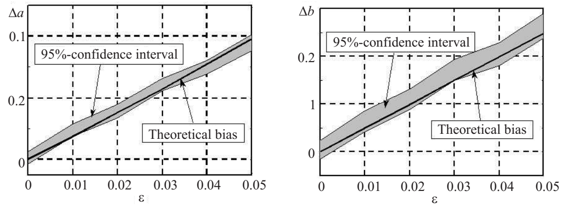


***Figure 3***: *The biases of the classical ML-estimator of the BLM parameters, which ignores the misclassifications: gray tubes –experimental 95% confidence intervals; solid lines– approximation via the expansion (15);* $\Delta a, \Delta b$ *–the relative bias,* $\varepsilon$*– the distortion level (*$\varepsilon_0$*=*$\varepsilon_1$*).*

*Experiment 2*. This experiment was aimed at the performance evaluation of the proposed robust estimator in the case of known distortion levels (Theorem 7). It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.03$, and the developed bias-corrected estimator was compared to the classical ML-estimator by assessing the biases, standard deviations, and histograms. As follows from the experiment results (Figure 4), the proposed estimation method allows essentially decreasing the bias of the $a, b$ estimates and leads to the similar standard deviations. In particular, the bias-corrected estimator yields the relative biases 2.3%, 1.1% for the parameters $a, b$ respectively against 23.1%, 8.2% obtained by applying the classical ML technique.

The identifiability of the model parameters $a, b$ and convergence of the numerical procedure (17) are determined by the properties of the $2 \times 2$ matrix of the second derivatives $J$ for the BLM log-likelihood function (which does not take into account the distortion model). Additional numerical research indicated that the determinant of this matrix was greater then $10^5$, while the corresponding condition number varied from 38.4 to 51.7. These results confirm both the identifiability and the robust performance of the bias-corrected estimator.

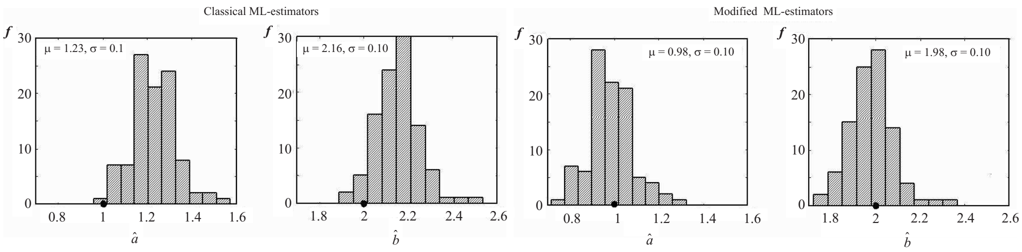


***Figure 4***: *Histograms of the classical and proposed estimators of the BLM parameters for known distortion levels:* $f$ *–empirical frequency,* $\mu$*– sample mean,* $\sigma$ *–sample standard deviation; the circles denote the true parameter values.*

*Experiment 3*. This experiment focused on the performance evaluation of the proposed robust estimator in the case of unknown distortion levels. It was assumed that $\varepsilon_0 = \varepsilon_1 = 0.03$, and the developed MLS-estimator was compared to the classical ML-estimator by assessing the biases and standard deviations. As follows from the experiment results (Table 5), the proposed estimation technique allows essentially decreasing the bias of the $a, b$ estimates, while the standard deviation is approximately the same in all cases. In particular, the MLS-estimator yields the relative biases 4.1%, 1.6% for the parameters $a, b$ respectively against 23.0%, 7.5% obtained by applying the classical ML technique.

To analyze the identifiability of the parameters $a, b, \varepsilon_0, \varepsilon_1$ and the convergence of the developed MLS estimation algorithm, there was examined the $4\times4$ matrix of the second derivatives for the log-likelihood function (18), which takes into account the distortion model. Its determinant was greater then $10^7$ that indicates the identifiability of all model parameters. But the corresponding condition number varied from $1.3{\cdot}10^3$ to $5.1\cdot10^4$ that explains relatively slow convergence of the optimization routine due to the ravine structure of the objective function. However, the computing time is acceptable for practical applications. These results confirm both the identifiability and the robust performance of the developed MLS-estimator.

***Table 5***: *Comparison of the classical and proposed estimators of the BLM for unknown distortion levels.*

| Parameter | $a$ (true value 1.0) | | $b$ (true value 2.0) | |
|---|---|---|---|---|
| Method | ML | MLS | ML | MLS |
| Mean | 1.23 | 1.04 | 2.15 | 2.03 |
| Standard deviation | 0.11 | 0.12 | 0.11 | 0.11 |

## 5 Robust forecasting for beta-mixed hierarchical models

**Robustness of the classical Bayes predictor.** First, let us analyze the robustness of the classical Bayes predictor (1), which incorporates the true values of the model parameters $\alpha_i^0,\ \beta_i^0$, assuming that the predictor input $x_i = \sum_{j=1}^{n_i} \tilde{B}_{ij}$ is contaminated by the distortions with known levels $\varepsilon_0,\ \ \varepsilon_1$ (here, the subscript $i$ denotes the index of the cluster, for which the forecast is performed). The following theorem evaluates the robustness of the classical Bayes predictor w.r.t. the distortion levels by assessing the increase of the mean square error of forecasting.

**Theorem 8** *If the classical Bayes predictor (1) uses the true model parameters $\alpha_i^0,\ \beta_i^0$, then the mean square error of the forecast, which is based in the misclassified responses, is expressed as*

$$\tilde{r}_i^2 = r_{0i}^2 + \frac{n_i(\beta_i^0\varepsilon_0 + \alpha_i^0\varepsilon_1) + n_i^{[2-]}((\beta_i^0)^{[2+]}\varepsilon_0^2 - 2\alpha_i^0\beta_i^0\varepsilon_0\varepsilon_1 + (\alpha_i^0)^{[2+]}\varepsilon_1^2)}{(\alpha_i^0 + \beta_i^0)^{[2+]}(\alpha_i^0 + \beta_i^0 + n_i)^2}. \quad (20)$$

*where $r_{0i}^2$ is the error in the non-distorted case* $(\varepsilon_0 = \varepsilon_1 = 0)$

$$r_{0i}^2 = \frac{\alpha_i^0\beta_i^0}{(\alpha_i^0 + \beta_i^0)^{[2+]}(\alpha_i^0 + \beta_i^0 + n_i)}.$$

Then, let us consider the case when the true values of the parameters $\alpha_i^0,\ \beta_i^0$ are unknown, so their estimates $\hat{\alpha}_i,\ \hat{\beta}_i$ (biased because of the distortions) are used for the prediction. The following theorem evaluates the robustness of the classical Bayes predictor in this case.

**Theorem 9** *If the classical Bayes predictor (1) uses the biased estimates $\hat{\alpha}_i,\ \hat{\beta}_i$ of the model parameters, then the mean square error of the forecast, which is based in the misclassified responses, is expressed as*

$$\tilde{r}_i^2 = \tilde{r}_{0i}^2 + \frac{n_i(\beta_i^0\eta_i^\beta\cdot\varepsilon_0 + \alpha_i^0\eta_i^\alpha\cdot\varepsilon_1)}{(\alpha_i^0+\beta_i^0)(\hat{\alpha}_i+\hat{\beta}_i+n_i)^2} + \frac{n_i^{[2-]}(\beta_i^{0[2+]}\cdot\varepsilon_0^2 - 2\alpha_i^0\beta_i^0\cdot\varepsilon_0\varepsilon_1 + \alpha_i^{0[2+]}\cdot\varepsilon_1^2)}{(\alpha_i^0+\beta_i^0)^{[2+]}(\hat{\alpha}_i+\hat{\beta}_i+n_i)^2}, \quad (21)$$

*where $\tilde{r}_{0i}^2$ is the error in the case of the non-distorted responses but the biased parameter esimates*

$$\tilde{r}_{0i}^2 = \frac{n_i\alpha_i^0\beta_i^0 + \beta_i^{0[2+]}\hat{\alpha}_i^2 - 2\alpha_i^0\beta_i^0\hat{\alpha}_i\hat{\beta}_i + \alpha_i^{0[2+]}\hat{\beta}_i^2}{(\alpha_i^0+\beta_i^0)^{[2+]}(\hat{\alpha}_i+\hat{\beta}_i+n_i)^2}, \quad (22)$$

*the coefficients $\eta_i^\alpha, \eta_i^\beta$ are*

$$\eta_i^\alpha = 1 + 2\frac{(\alpha_i^0+1)\hat{\beta}_i - (\hat{\alpha}_i+1)\beta_i^0}{\alpha_i^0+\beta_i^0+1}, \qquad \eta_i^\beta = 1 + 2\frac{(\beta_i^0+1)\hat{\alpha}_i - (\hat{\beta}_i+1)\alpha_i^0}{\alpha_i^0+\beta_i^0+1}, \quad (23)$$

*and the ascending and descending factorials are denoted as* $v^{[2+]} = v(v+1),\ v^{[2-]} = v(v-1)$.

As follows from these theorems, the classical Bayes predictor loses its optimality under the distortions (in the sense of the mean square error of forecasting). Expressions (20), (21) allow assessing the sensitivity of the classical predictor to the misclassifications (3). Let us propose now the robust predictor that takes into account the distortion model (2).

**Robust prediction under distortions.** Since the results from the previous sections allow obtaining the unbiased estimates of the model parameters as well as the probability distribution of the misclassified responses, there can be derived the optimal predictor that minimizes the effect of the misclassifications in the forecast input data. This predictor is defined in the following theorem.

**Theorem 10** *The optimal Bayes predictor, which takes into account the distortion model (2), (3), is expressed as the weighted sum*

$$\hat{p}_i(x) = \mathbf{E}\{p_i|x,\varepsilon_0,\varepsilon_1\} = \sum_{r=0}^{n_i}\vartheta_{xr}^i\cdot\frac{\alpha_i^0+r}{\alpha_i^0+\beta_i^0+n_i}, \quad (24)$$

*where $x$ is the sum of the distorted binary observations for the $i$-th cluster, and the weighting coefficients $\vartheta^i_{xr}$ are computed from*

$$\vartheta^i_{xr} = \binom{n_i}{r} w^i_{xr}\, B(\alpha^0_i + r, \beta^0_i + n_i - r) \cdot \left( \sum_{l=0}^{n_i} \binom{n_i}{l} w^i_{xl}\, B(\alpha^0_i + l, \beta^0_i + n_i - l) \right)^{-1} \qquad (25)$$

*using expressions for $w^i_{sl}$ given in Theorem 1.*

It can also be proved that the corresponding mean square error of forecasting is computed as

$$r^2(\hat{p}_i) = \frac{(\alpha^0_i)^{[2+]}}{(\alpha^0_i + \beta^0_i)^{[2+]}} - \sum_{x=0}^{n_i} \left( \left( \sum_{r=0}^{n_i} \vartheta^i_{xr} \frac{\alpha^0_i + r}{\alpha^0_i + \beta^0_i + n_i} \right)^2 \cdot \sum_{r=0}^{n_i} w^i_{xr} \binom{n_i}{x} \frac{(a^0_i)^{[r+]} (\beta^0_i)^{[(n_i - r)+]}}{(\alpha^0_i + \beta^0_i)^{[n_i+]}} \right), \qquad (26)$$

and the p.d.f. of this forecast is

$$f_{p_i}(p|x, \varepsilon_0, \varepsilon_1) = \sum_{r=0}^{n_i} \vartheta^i_{xr} \cdot B(\alpha^0_i + r, \beta^0_i + n_i - r)^{-1}\, p^{\alpha^0_i + r - 1} (1 - p)^{\beta^0_i + n_i - r - 1}. \qquad (27)$$

As follows from expressions (24), (27), the proposed predictor is a weighted sum of the classical predictors for the beta-hierarchical model with shifted parameters. Also, the expression for the weights $\vartheta^i_{xr}$ are based on the Bayes formula, and $\vartheta^i_{xr}$ can be treated as the posteriori probability that the distorted value $x$ was originated from the sum of the non-distorted binary observations $r$ (in contrast, $w^i_{xr}$ are the corresponding a priori probabilities). Since for the weight matrix $(\vartheta^i_{xr})$, there can be obtained the asymptotic expansion similar to (5), it is prudent to derive an approximate expression for the proposed predictor (24), which is valid for small values of $\varepsilon_0, \varepsilon_1$.

**Robust prediction for small distortion levels.** If values of $\varepsilon_0, \varepsilon_1$ are small, then the sums in expressions (24), (25) can be reduced to three terms by eliminating the weighting coefficients other then $\vartheta^i_{x,x-1}$, $\vartheta^i_{x,x}$, $\vartheta^i_{x,x+1}$ for (24) and $w^i_{x,x-1}$, $w^i_{x,x}$, $w^i_{x,x+1}$ for (25). Then the robust predictor can be expressed as the classical Bayes predictor multiplied by the correction factor

$$\hat{p}_i(x) = \frac{\alpha^0_i + x}{\alpha^0_i + \beta^0_i + n_i} \cdot \frac{1 + \gamma_0 \cdot \varepsilon_0 - \gamma_1 \cdot \varepsilon_1}{1 + \xi_0 \cdot \varepsilon_0 - \xi_1 \cdot \varepsilon_1} + o(\varepsilon_0, \varepsilon_1), \qquad (28)$$

where

$$\gamma_0 = \frac{(\alpha^0_i + \beta^0_i)x - \alpha^0_i n_i}{\alpha^0_i + x}, \qquad \gamma_1 = \frac{(\alpha^0_i + \beta^0_i)x - \alpha^0_{i,+} n_i}{\beta^0_{i,-} + n_i - x},$$

$$\xi_0 = \frac{(\alpha^0_{i,-} + \beta^0_i)x - \alpha^0_{i,-} n_i}{\alpha^0_{i,-} + x}, \qquad \xi_1 = \frac{(\alpha^0_i + \beta^0_{i,-})x - \alpha^0_i n_i}{\beta^0_{i,-} + n_i - x},$$

and $\alpha^0_{i,-} = \alpha^0_i - 1$, $\beta^0_{i,-} = \beta^0_i - 1$, $\alpha^0_{i,+} = \alpha^0_i + 1$. Expression (28) allows essentially simplifying the complexity of the robust forecasting algorithm and can be used in practical applications, for which the computing time is crucial.

**Computer simulation.** To demonstrate the robust performance of the developed forecasting technique, the following computer simulation was done. There was considered the beta-binomial model, and it was assumed that the true values of the model parameters were $\alpha^0 = 0.5$, $\beta^0 = 9.5$, $n = 10$. For this $\alpha^0, \beta^0$, there were generated $k = 1000$ realizations of the beta random variable $p_1, p_2, \ldots, p_k$ (the corresponding mean value was $\bar{p} = 0.05$). Then, for each cluster with the success probability $p_i$, a random Bernoulli sample of size $n$ was obtained. Next, every sample was distorted using the expression (2) for $\varepsilon_0 = \varepsilon_1 \in [\,0;\ 0.05\,]$ varying with the step 0.01. Using these data, the ML- and MLS-estimates of the $\alpha, \beta$ parameters were computed. For each cluster, two types of the forecast was done: (i) the classical prediction (1) based on the ML-estimates, and (ii) the proposed prediction (24) based on the MLS-estimates. Finally, for every distortion level, the 95%-confidence intervals of the mean square error of forecasting were computed for the both predictors (assuming that the errors follow the normal distribution).

As follows from the experiment results (Figure 5), the developed prediction technique based on the proposed MLS-estimation algorithm ensures essentially lower mean square error of forecasting when compared to the classical estimation and prediction methods.

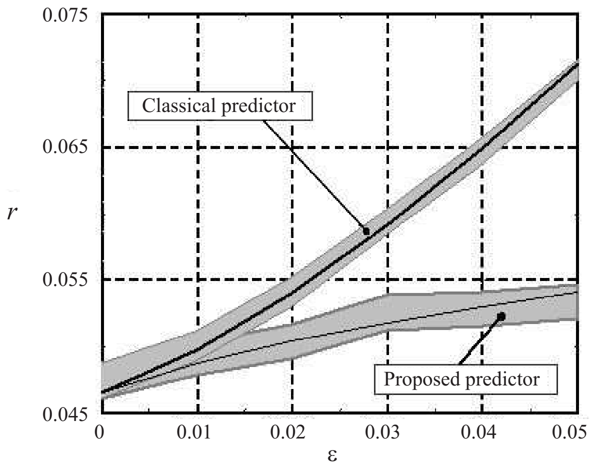


***Figure 5***: *Comparison of the classical and proposed predictors: gray tubes – experimental 95% confidence intervals, solid lines– theoretical mean square errors of forecasting; r – mean square error, ε – distortion level ($\varepsilon_0$=$\varepsilon_1$).*

For example, for the distortion levels $\varepsilon_0 = \varepsilon_1 = 0.05$, the classical procedures lead to the error $r = 0.071$ against $r_0 = 0.047$ for the non-distorted case, while the proposed robust methods ensure the error $r = 0.054$ (note that Figure 5 shows $r$ against $\varepsilon$, while the above expressions are given for $r^2$). Hence, for the average response probability $\bar{p} = 0.05$, the increment of the forecast error $\Delta r = r - r_0$ reduces from 0.026 to 0.007. It should be stressed that the increment $\Delta r$ caused the misclassifications can not be compensated completely (as follows from the Bayes forecasting theory), but the obtained value $r = 0.054$ is the lowest for these model parameters and distortion levels. These results confirm the robust performance of the developed estimation and forecasting techniques.

## 6 Application example

The developed methods of robust estimation and prediction were used for forecasting TV audience behaviour. This problem arises in mediaplanning (Sissors and Lincoln 1994), which focuses on optimizing of advertising schedules taking into account the target consumer groups (defined by age, sex, income, etc.) and budget constraints. For this application area, statistical forecasting of future audience behaviour using records from the past is a key issue, since it defines efficiency of the advertising spending.

**Grouped binary data in mediaplanning.** In TV mediaplanning, the binary responses arise as a result of exposing advertising commercials to a part of TV audience (the representative sample of the target group) during predefined TV breaks, where 1 means that a person saw the commercial and vice versa. These data are registered by special electronic devices (people-meters) and are grouped in a natural way with respect to every person and break type (defined by week day, day time, adjoining program genres, etc.).

The misclassifications that may contaminate these data are caused by improper use of the people-meters, which automatically register a TV channel being viewed, but require manual registration of household members watching the TV. It is obvious that there exists a small probability of using a wrong registration button that leads to distortions of the recorded observations. The statistical properties of the viewing data are traditionally described by the beta-binomial model (Danaher 1992), while the misclassification effect is usually ignored.

In frames of the paper notation, the TV viewing data may be interpreted as follows: $\tilde{B}_{ij}$ is the $i$-th person response to $j$-th commercial break of the certain type, $k$ is the number of persons in a target group, and $n_i$ is the number of the breaks that the $i$-th person was exposed to. It assumed that the target group and break type uniquely define the covariates $Z_i$, and each person's viewing behaviour for this break type is described by the success probability $p_i$ that follows the beta distribution with the parameter $\alpha_i^0, \beta_i^0$.

For the case studies below, there were examined two data sets for one of the German TV channels for the year 2000. The first of them focuses on improving the model adequacy, while the second one deals with increasing the forecasting accuracy.

**Viewing data modelling.** To demonstrate the advantages of the developed distorted beta-binomial model (DBBM), which takes into account the misclassifications and employs the proposed robust estimation techniques, there were considered the TV viewing data for eleven commercial breaks ($n$ = 11) corresponding to "World News" showed on Saturday prime time. There were investigated six target groups with different sex (M,W) and age (14-29, 30-49, 50+) with size $k$ varying from 1025 to 2488.

The results of the model adequacy analysis are presented in Table 6, which shows that the proposed DBBM and the relevant robust estimation algorithms significantly

***Table 6***: *Adequacy analysis of the classical (BBM) and the proposed (DBBM) models for describing the TV audience behaviour using Pearson's $\chi^2$ goodness-of-fit statistics.*

| Target group | M 14-29 | M 30-49 | M 50+ | W 14-29 | W 30-49 | W 50+ |
|---|---|---|---|---|---|---|
| Data characteristics | | | | | | |
| Group size, $k$ | 1137 | 2011 | 2281 | 1025 | 2084 | 2488 |
| Sample mean | $1.2\cdot10^{-2}$ | $4.9\cdot10^{-2}$ | $3.6\cdot10^{-2}$ | $2.1\cdot10^{-2}$ | $3.8\cdot10^{-2}$ | $4.7\cdot10^{-2}$ |
| Overdispersion | 1.66 | 3.33 | 2.75 | 2.05 | 2.54 | 3.27 |
| Classical beta-binomial model (MM-estimator) | | | | | | |
| $p$-value | 0.41 | 0.96 | 0.01 | 0.82 | 0.10 | 0.05 |
| $\chi^2$-statistics | 9.30 | 3.14 | 21.7 | 5.16 | 14.6 | 17.1 |
| Parameter $\alpha$ | 0.17 | 0.16 | 0.17 | 0.18 | 0.21 | 0.16 |
| Parameter $\beta$ | 13.9 | 3.14 | 4.56 | 8.33 | 5.27 | 3.25 |
| Classical beta-binomial model (ML-estimator) | | | | | | |
| $p$-value | 0.45 | 0.97 | 0.02 | 0.80 | 0.10 | 0.05 |
| $\chi^2$-statistics | 8.88 | 2.87 | 20.0 | 5.43 | 14.6 | 16.9 |
| Parameter $\alpha$ | 0.17 | 0.17 | 0.19 | 0.19 | 0.24 | 0.17 |
| Parameter $\beta$ | 13.5 | 6.55 | 5.08 | 8.81 | 5.78 | 3.47 |
| Distorted beta-binomial model (MMS-estimator) | | | | | | |
| $p$-value | 0.28 | 0.99 | 0.32 | 0.89 | 0.84 | 0.14 |
| $\chi^2$-statistics | 10.9 | 1.00 | 10.4 | 4.35 | 4.99 | 13.5 |
| Parameter $\alpha$ | 0.09 | 0.13 | 0.14 | 0.13 | 0.15 | 0.15 |
| Parameter $\beta$ | 9.57 | 5.58 | 3.88 | 7.07 | 4.00 | 3.14 |
| Distortion level $\varepsilon_0$ | 0.003 | 0.002 | 0.003 | 0.003 | 0.006 | 0.001 |
| Distortion level $\varepsilon_1$ | 0.060 | 0.000 | 0.042 | 0.000 | 0.060 | 0.004 |
| Distorted beta-binomial model (MLS-estimator) | | | | | | |
| $p$-value | 0.63 | 0.98 | 0.77 | 0.88 | 0.83 | 0.50 |
| $\chi^2$-statistics | 7.11 | 2.46 | 5.67 | 4.40 | 5.04 | 8.34 |
| Parameter $\alpha$ | 0.11 | 0.15 | 0.10 | 0.15 | 0.15 | 0.11 |
| Parameter $\beta$ | 9.12 | 6.39 | 3.14 | 7.51 | 4.31 | 2.36 |
| Distortion level $\varepsilon_0$ | 0.001 | 0.002 | 0.007 | 0.002 | 0.006 | 0.005 |
| Distortion level $\varepsilon_1$ | 0.048 | 0.015 | 0.068 | 0.008 | 0.020 | 0.111 |

increase the modelling accuracy. For example, for the target group M 50+ (men of age 50 and older), the classical BBM yields the $p$-values 0.01 for the MM-estimator and 0.02 for the ML-estimator, while the proposed DBBM ensures values 0.32 and 0.77 for the MMS and MLS estimators respectively. This confirms the applicability of the paper results to the modelling of the TV audience behaviour.

**Forecasting of audience behaviour.** To illustrate the accuracy of the developed forecasting technique, there were considered $N_z = 31$ commercial breaks of different types exposed in December 2000 for the target group W 50+ in the frames of a single adverting campaign. Based on the past data for the similar breaks (for three preceding months, September – November, 2000), there were obtained the viewing behaviour models based on the proposed DBBD distribution. Then, using the proposed prediction method, for all persons and all breaks, there were generated the forecasts $\pi_{iz}$ (the probability that the $i$-th person watched the break of type $z$). Similar forecasts were also obtained for the classical model based on the BBD.

The accuracy for the obtained forecast was evaluated using the specific performance measures adopted in mediaplanning, the *Reach* and *GRP* (Danaher 1992). The first of them, *Reach*, describes the audience fraction (within the target group), which have seen the advertising commercial at least once during the whole advertising campaign:

$$Reach = k^{-1} \sum_{i=1}^{k} \left( 1 - \prod_{z=1}^{N_z} (1 - \pi_{iz}) \right).$$

The second performance measure, *GRP* (Gross Rating Points), defines the sum of the above fractions throughout the campaign (without considering the audience duplication):

$$GRP = k^{-1} \sum_{i=1}^{k} \sum_{z=1}^{N_z} \pi_{iz} \ .$$

Using these expressions, there were obtained the *Reach–GRP* curves via considering smaller advertising campaigns composed of the considered breaks (with break number from 1 to $N_z$). In practice, such curves are the primary tool for media-planners who use them for assessing the economical efficiency of adding extra break to the campaign.

Figure 6 compares the *Reach–GRP* curves for the BBM and DBBM-based forecasts with the real data curve calculated using the December 2000 records. As follows from the figure, the proposed forecasting technique ensures much more accurate approximation of the *Reach–GRP* relation than the classical BBM method. In particular, the maximum relative error of the *Reach–GRP* approximation using the BBM-based forecast is about 21%, while the proposed DBBM-based technique ensures the relative error less then 4.2%. This confirms the practical value of our results.

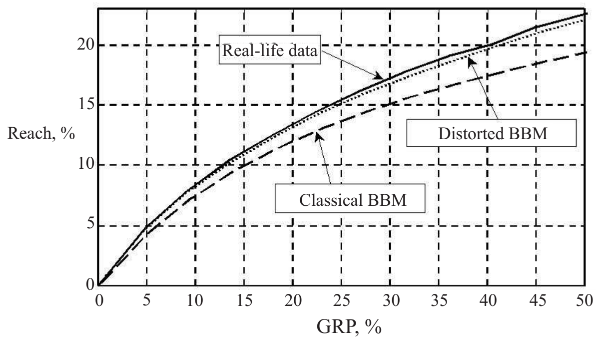


***Figure 6***: *Comparison of the Reach–GRP curves based on the classical (BBM) and the proposed (DBBM) models against the curve obtained from the real data.*

## 7 Conclusion

The paper proposes new robust estimation and forecasting techniques for the grouped binary data in the case of response misclassifications caused by stochastic additive distortions. It is assumed that the data are described by the beta-binomial or the beta-logistic model that belong to the class of the beta-mixed hierarchical ones. For these models, it is examined the effect of ignoring the misclassifications and there are obtained expressions for the biases of the method-of-moments and maximum likelihood estimators, as well as expressions for the increase in the mean square error for the Bayes predictor. These expressions allow assessing the sensitivity of the classical techniques w.r.t. the distortion levels and decide on their applicability in practice.

To minimize the misclassification effects, there were developed new consistent estimators and a new Bayes predictor, which take into account the distortion model. There were considered two cases (of known and unknown distortion levels), for which explicit expressions and numerical algorithms were proposed that allow constructing the small-sensitive estimators of the model parameters and the small-sensitive forecasting procedures. The robustness of the developed techniques was verified by computer simulations, and the practical value was confirmed by a real-life case study. The proposed algorithms were implemented as a MATLAB toolbox.
Future work will deal with the minimax robust estimation and forecasting for the case of known upper and lower bounds of the distortion levels, and also with the problem of small sample performance for the developed methods.

**Acknowledgments** This research was partly supported by the ISTC project B-705 and INTAS grant YSF 03-55-869. The authors thank Prof. Vladimir Zaiats for the financial support at the Barcelona Conference on Asymptotic Statistics, 2003. The authors are also grateful to Omega Software GmbH for the application example data and to the Reviewers whose valuable comments improved the paper.

## Mathematical Appendix

**Basic notation.** $\mathbf{P}\{.\}$ is the probability of a random event, $\mathbf{E}\{.\}$ is the mathematical expectation of a random variable, $\mathbf{V}\{.\}$ is the variance of a random variable, $y^{[z-]} = y(y-1)\dots(y-z+1), y^{[z+]} = y(y+1)\dots(y+z-1), y \in R, z \in N$ are the incomplete factorials, $C_n^r = \binom{n}{r}$ is the binomial coefficient. Definition of $O_P$: for two random sequences $Y_n, Z_n$, $Y_n = O_P(Z_n)$ means that $\forall \epsilon > 0 \quad \exists\, k_\epsilon, N_\epsilon$ that $0 < k_\epsilon < +\infty$, $0 < N_\epsilon < +\infty$ and for $n > N_\epsilon$, $\mathbf{P}\{|Y_n/Z_n| < k_\epsilon\} > 1 - \epsilon$.

*Proof of Theorem 1.* Let $r$ be a realization of the DBBD random variable. Denote by $\{H_{rs}\}, r = 0, 1, \dots, n$, a partition complete set of disjoint events, where $H_{rs}$ means that the distorted value $r$ was obtained via the distortions (2) from the original positive responses count $s$. Then using the total probability formula

$$P_r(\varepsilon_0, \varepsilon_1) = \sum_{s=0}^{n} \mathbf{P}\{H_{rs}\} \cdot P_s^0, r = 0, 1, \dots, n.$$

To find the probability $\mathbf{P}\{H_{rs}\}$, denote by $z_0, z_1$ the number of the distorted zeros and ones in the original data. Then combinatorics yields to the following expression

$$w_{rs}(\varepsilon_0, \varepsilon_1) = \mathbf{P}\{H_{rs}\} = \sum_{z_0, z_1} C_n^{z_1} \varepsilon_0^{z_0} (1 - \varepsilon_0)^{n-s-z_0} \varepsilon_1^{z_1} (1 - \varepsilon_1)^{s-z_1}, \quad s - z_1 + z_0 = r.$$

Denoting $l = s + z_0 = r + z_1$ leads to $l \geq r$, $l \leq s + r$, $l \geq s$, $l \leq n$, which is equivalent to $\max(s, r) \leq l \leq \min(n, s + r)$, that proves the theorem. □

**Remark.** The standard approach for investigating the properties of the estimators that are fitted to the misspecified model is based on the results of White (1982) that involve Kullback-Leibler divergence. For Theorems 2, 3, 6, the authors employ a different approach that allows using the specific DBBD properties to obtain elegant proofs. However, one can check that using the Kullback-Leibler divergence leads to the exactly the same results.

*Proof of Theorem 2.* The classical MM-estimator of the BBM parameters is expressed as (Johnson *et al.* 1996):

$$\tilde{\alpha}_{MM} = \frac{(n - \overline{x} - s^2/\overline{x})\overline{x}}{(s^2/\overline{x} + \overline{x}/n - 1)n}, \quad \tilde{\beta}_{MM} = \frac{(n - \overline{x} - s^2/\overline{x})(n - \overline{x})}{(s^2/\overline{x} + \overline{x}/n - 1)n}, \tag{29}$$

where $\overline{x}$ is the sample average and $s^2$ is the sample variance. Let $m(\varepsilon_0, \varepsilon_1)$, $d(\varepsilon_0, \varepsilon_1)$ be the mean and variance of the DBBD with the parameters $n, \alpha^0, \beta^0, \varepsilon_0, \varepsilon_1$ (see Theorem 1). Since $\overline{x}$, $s^2$ are unbiased and consistent estimators, and $\mathbf{V}\{\overline{x}\} = d(\varepsilon_0, \varepsilon_1)/k$,

$\mathbf{V}\{s^2\} = O(1/k)$ (Ivchenko and Medvedev 1984), then $\overline{x} = m(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k})$, $s^2 = d(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k})$. Using these expressions together with the properties of $O_P(.)$ to modify (29), we get

$$\tilde{\alpha}_{MM}(\varepsilon_0, \varepsilon_1) = \frac{(n - m(\varepsilon_0, \varepsilon_1) - d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1))\, m(\varepsilon_0, \varepsilon_1)}{(d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1) + m(\varepsilon_0, \varepsilon_1)/n - 1)n} + O_P(1/\sqrt{k}),$$

$$\tilde{\beta}_{MM}(\varepsilon_0, \varepsilon_1) = \frac{(n - m(\varepsilon_0, \varepsilon_1) - d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1))\, (n - m(\varepsilon_0, \varepsilon_1))}{(d(\varepsilon_0, \varepsilon_1)/m(\varepsilon_0, \varepsilon_1) + m(\varepsilon_0, \varepsilon_1)/n - 1)\, n} + O_P(1/\sqrt{k}).$$

Employing the expressions for $m(\varepsilon_0, \varepsilon_1)$, $d(\varepsilon_0, \varepsilon_1)$ and the linear term of the Taylor expansion with the Peano remainder for the above functions of $\varepsilon_0, \varepsilon_1$ proves the theorem. □

**Expressions for Theorem 3.** In the theorem statement, the following notation is used:

$$P^0_{0,s} = P^0_s(\alpha^0, \beta^0), \quad P^{\Sigma}_s(\varepsilon_0, \varepsilon_1) = \sum_{r=0}^{s} P_r(\alpha^0, \beta^0, \varepsilon_0, \varepsilon_1), \quad s = 0, 1, \ldots, n,$$

$$S_\alpha = \sum_{s=0}^{n-1} \frac{1 - P^{\Sigma}_s(0,0)}{(\alpha^0 + s)^2}, \quad S_\beta = \sum_{s=0}^{n-1} \frac{P^{\Sigma}_s(0,0)}{(\beta^0 + n - s - 1)^2}, \quad S_{\alpha\beta} = \sum_{s=0}^{n-1} \frac{1}{(\alpha^0 + \beta^0 + s)^2},$$

$$S_{\alpha p} = -\sum_{s=0}^{n-1} \frac{(n-s)P^0_{0,s}}{\alpha^0 + s}, \; S^+_{\alpha p} = \sum_{s=0}^{n-1} \frac{(s+1)P^0_{0,s+1}}{\alpha^0 + s}, \; S_{\beta p} = \sum_{s=0}^{n-1} \frac{(n-s)P^0_{0,s}}{\beta^0 + n - s - 1},$$

$$S^+_{\beta p} = -\sum_{s=0}^{n-1} \frac{(s+1)P^0_{0,s+1}}{\beta^0 + n - s - 1},$$

$$H = \{H_{ij}\}_{2\times 2}, \quad G = \{G_{ij}\}_{2\times 2}, \quad H_{11} = S_{\alpha\beta} - S_\alpha,$$

$$H_{12} = H_{21} = S_{\alpha\beta}, \; H_{22} = S_{\alpha\beta} - S_\beta, \; G_{11} = S_{\alpha p}, \; G_{12} = S^+_{\alpha p} = S^+_{\beta p} = G_{21}, \; G_{22} = S_{\beta p}.$$

*Proof of Theorem 3.* The ML-estimator for the BBM is defined as a solution of the following system of two equations (Johnson *et al.* 1996)

$$\sum_{r=0}^{n-1} \frac{k - F_r}{\alpha + r} - \sum_{i=0}^{n-1} \frac{k}{\alpha + \beta + r} = 0, \quad \sum_{r=0}^{n-1} \frac{F_r}{\beta + n - r - 1} - \sum_{r=0}^{n-1} \frac{k}{\alpha + \beta + r} = 0, \qquad (30)$$

where $F_r = f_0 + f_1 + \ldots + f_r$, and $\{f_s\}$ are the empirical frequencies. The system has a single solution that maximizes the likelihood function (Johnson *et al.* 1996). By definition, the frequencies are the binomial random variables with the parameters $k$, $P_s(\alpha^0, \beta^0, \varepsilon_0, \varepsilon_1)$. Since for a discrete probability distribution, the relative frequencies $\widetilde{f_s} = f_s/k$ are unbiased and consistent estimators of the corresponding theoretical probabilities, and

$\mathbf{V}\{\widetilde{f_s}\} = P_s(\alpha^0, \beta^0, \varepsilon_0, \varepsilon_1)(1 - P_s(\alpha^0, \beta^0, \varepsilon_0, \varepsilon_1))/k$, then $\widetilde{f_s} = f_s/k = P_s(\alpha^0, \beta^0, \varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k})$, $s = 0, 1, ..., n$. As a result, the system (30) can be expressed as

$$\sum_{r=0}^{n-1} \frac{1 - P_r^{\Sigma}(\varepsilon_0, \varepsilon_1)}{\alpha + r} - \sum_{r=0}^{n-1} \frac{1}{\alpha + \beta + r} + O_P\left(\frac{1}{\sqrt{k}}\right) = 0,$$

$$\sum_{r=0}^{n-1} \frac{P_r^{\Sigma}(\varepsilon_0, \varepsilon_1)}{\beta + n - r - 1} - \sum_{r=0}^{n-1} \frac{1}{\alpha + \beta + r} + O_P\left(\frac{1}{\sqrt{k}}\right) = 0.$$

Let us linearize the obtained system by $\varepsilon_0, \varepsilon_1$ in the neighborhood of the point $(\alpha^0, \beta^0, 0, 0)$, then

$$A_\alpha^0 \Delta\widetilde{\alpha}_{ML}(\varepsilon_0, \varepsilon_1) + A_\beta^0 \Delta\widetilde{\beta}_{ML}(\varepsilon_0, \varepsilon_1) + A_{\varepsilon_0}^0 \varepsilon_0 + A_{\varepsilon_1}^0 \varepsilon_1 + o(\varepsilon_0) + o(\varepsilon_1) + O_P(1/\sqrt{k}) = 0,$$

$$B_\alpha^0 \Delta\widetilde{\alpha}_{ML}(\varepsilon_0, \varepsilon_1) + B_\beta^0 \Delta\widetilde{\beta}_{ML}(\varepsilon_0, \varepsilon_1) + B_{\varepsilon_0}^0 \varepsilon_0 + B_{\varepsilon_1}^0 \varepsilon_1 + o(\varepsilon_0) + o(\varepsilon_1) + O_P(1/\sqrt{k}) = 0,$$

where the coefficients are the corresponding derivatives. Expressing the $\Delta\widetilde{\alpha}_{ML}(\varepsilon_0, \varepsilon_1)$, $\Delta\widetilde{\beta}_{ML}(\varepsilon_0, \varepsilon_1)$ in terms of $\varepsilon_0, \varepsilon_1$ from this system proves the theorem. □

*Proof of Theorem 4*. Using Theorem 1, one can show that the MM-estimator of the BBM parameters $\alpha, \beta$ that takes into account the distortions model (2) is defined as a solution of the following system of two equations

$$m_1^* = n\frac{\alpha}{\alpha + \beta} + n\frac{\beta}{\alpha + \beta} \cdot \varepsilon_0 - n\frac{\alpha}{\alpha + \beta} \cdot \varepsilon_1, \tag{31}$$

$$m_2^* = m_1^* + n^{[2-]} \cdot \frac{\alpha^{[2+]} + \beta^{[2+]}\varepsilon_0^2 + \alpha^{[2+]}\varepsilon_1^2 - 2\alpha\beta\varepsilon_0 - \alpha^{[2+]}\varepsilon_1 - 2\alpha\beta\varepsilon_0\varepsilon_1}{(\alpha + \beta)^{[2+]}}. \tag{32}$$

Using the substitution

$$u = \frac{\alpha}{\alpha + \beta}, \quad v = \frac{\alpha + 1}{\alpha + \beta + 1}, \quad \alpha = \frac{u(1 - v)}{v - u}, \quad \beta = \frac{(1 - v)(1 - u)}{v - u},$$

transforms the above system into

$$m_1^* = n(u + (1 - u)\varepsilon_0 - u\varepsilon_1), \quad m_2^* = m_1^* + n(n - 1)(vu(1 - \varepsilon_0 - \varepsilon_1) + \varepsilon_0^2 + 2u\varepsilon_0(1 - \varepsilon_0 - \varepsilon_1)).$$

Solving this system with respect to $u$, $v$ and changing the variables back to $\alpha, \beta$ proves the theorem. □

*Proof of Theorem 5*. The empirical probabilities vector $\hat{P}(\varepsilon_0, \varepsilon_1)$ satisfies the following asymptotic expression (see the proof of Theorem 3): $\hat{P}_r(\varepsilon_0, \varepsilon_1) = P_r(\varepsilon_0, \varepsilon_1) + O_P(1/\sqrt{k})$,

$r = 0, 1, \ldots, n$. Using the result of Theorem 1, one gets $\hat{P}_r(\varepsilon_0, \varepsilon_1) = W(\varepsilon_0, \varepsilon_1) \cdot P^0 + O_P(1/\sqrt{k})$. Using the properties of $O_P(1/\sqrt{k})$ and the notation (10) concludes the proof. □

**MMS-estimator of BBM parameters.** The Jacobi matrix $J_0^c$ for the iterative procedure (13) is calculated as $J_0^c = H \cdot G + S$, where

$$H_{11} = n^{[3-]} \frac{\alpha^{[3+]}}{(\alpha+\beta)^{[3+]}} \sum_{i=0}^{2} \left( \frac{1}{\alpha + i} - \frac{1}{\alpha+\beta+i} \right),$$

$$H_{12} = n^{[3-]} \frac{\alpha^{[3+]}}{(\alpha+\beta)^{[3+]}} \sum_{i=0}^{2} \left( \frac{-1}{\alpha+\beta+i} \right),$$

$$H_{21} = n^{[4-]} \frac{\alpha^{[4+]}}{(\alpha+\beta)^{[4+]}} \sum_{i=0}^{3} \left( \frac{1}{\alpha + i} - \frac{1}{\alpha+\beta+i} \right) + 6 \cdot H_{11},$$

$$H_{22} = n^{[4-]} \frac{\alpha^{[4+]}}{(\alpha+\beta)^{[4+]}} \sum_{i=0}^{3} \left( \frac{-1}{\alpha+\beta+i} \right) + 6 \cdot H_{12},$$

$$G_{11} = -(\alpha + 2\beta + 1), \quad G_{12} = -\alpha(\alpha+1)/\beta,$$
$$G_{21} = -\beta(\beta+1)/\alpha, \quad G_{22} = -\left(2\alpha^0 + \beta^0 + 1\right),$$
$$S_{11} = 3n^{[3-]} \frac{\alpha^{[2+]}\beta}{(\alpha+\beta)^{[3+]}}, \qquad S_{12} = 3n^{[3-]} \frac{\alpha^{[3+]}}{(\alpha+\beta)^{[3+]}},$$
$$S_{21} = 14n^{[4-]} \frac{\alpha^{[3+]}\beta}{(\alpha+\beta)^{[4+]}} + 6 \cdot S_{11}, \qquad S_{22} = 4n^{[4-]} \frac{\alpha^{[4+]}\beta}{(\alpha+\beta)^{[4+]}} + 6 \cdot S_{12}.$$

**MLS-estimator of BBM parameters.** The partial derivatives of the log-likelihood function $l(\alpha, \beta, \varepsilon_0, \varepsilon_1)$ are computed as

$$\frac{\partial l}{\partial \alpha} = \sum_{r=0}^{n} \left( f_r \sum_{i=0}^{n} w_{ri}(\varepsilon_0, \varepsilon_1) \cdot \frac{\partial P_i^0(\alpha, \beta)/\partial \alpha}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

$$\frac{\partial l}{\partial \beta} = \sum_{r=0}^{n} \left( f_r \sum_{i=0}^{n} w_{ri}(\varepsilon_0, \varepsilon_1) \cdot \frac{\partial P_i^0(\alpha, \beta)/\partial \beta}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

$$\frac{\partial l}{\partial \varepsilon_0} = \sum_{r=0}^{n} \left( f_r \sum_{i=0}^{n} \frac{\partial w_{ri}(\varepsilon_0, \varepsilon_1)}{\partial \varepsilon_0} \cdot \frac{P_i^0(\alpha, \beta)}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

$$\frac{\partial l}{\partial \varepsilon_1} = \sum_{r=0}^{n} \left( f_r \sum_{i=0}^{n} \frac{\partial w_{ri}(\varepsilon_0, \varepsilon_1)}{\partial \varepsilon_1} \cdot \frac{P_i^0(\alpha, \beta)}{P_r(\alpha, \beta, \varepsilon_0, \varepsilon_1)} \right),$$

where

$$\frac{\partial P_i^0}{\partial \alpha} = P_i^0(\alpha,\beta) \cdot \left( \sum_{j=0}^{i-1} \frac{1}{\alpha + j} - \sum_{j=0}^{n-1} \frac{1}{\alpha + \beta + j} \right),$$

$$\frac{\partial P_i^0}{\partial \beta} = P_i^0(\alpha,\beta) \cdot \left( \sum_{j=0}^{n-i-1} \frac{1}{\beta + j} - \sum_{j=0}^{n-1} \frac{1}{\alpha + \beta + j} \right),$$

$$\frac{\partial w_{ri}}{\partial \varepsilon_0} = \sum_{l=max(i,r)}^{min(n,i+r)} C_i^{l-r} C_{n-i}^{l-i} \left( (l-i)\varepsilon_0^{l-i-1}(1-\varepsilon_0)^{n-l} - (n-l)\varepsilon_0^{l-i}(1-\varepsilon_0)^{n-l-1} \right) \varepsilon_1^{l-r}(1-\varepsilon_1)^{i+r-l},$$

$$\frac{\partial w_{ri}}{\partial \varepsilon_1} = \sum_{l=max(i,r)}^{min(n,i+r)} C_i^{l-r} C_{n-i}^{l-i} \varepsilon_0^{l-i}(1-\varepsilon_0)^{n-l} \left( (l-r)\varepsilon_1^{l-r-1}(1-\varepsilon_1)^{i+r-l} - (i+r-l)\varepsilon_1^{l-r}(1-\varepsilon_1)^{i+r-l-1} \right).$$

**Expressions for Theorem 6.** In the theorem statement, the following notation is used:

$$H_{ls} = \sum_{q=1}^{d} \vartheta_{ql}\vartheta_{qs}\breve{\alpha}_q^0 \left( \sum_{j=0}^{\breve{n}_q-1} \left( \frac{1-\pi_j^q}{\breve{\alpha}_q^0 + j} - \frac{1}{\breve{\alpha}_q^0 + \breve{\beta}_q^0 + j} \right) - \sum_{j=0}^{\breve{n}_q-1} \left( \frac{1-\pi_j^q}{(\breve{\alpha}_q^0 + j)^2} - \frac{1}{(\breve{\alpha}_q^0 + \breve{\beta}_q^0 + j)^2} \right) \breve{\alpha}_q^0 \right),$$

$$l, s = 1, \ldots, m,$$

$$H_{ls} = \sum_{q=1}^{d} \vartheta_{ql}\vartheta_{qs}\breve{\beta}_q^0 \left( \sum_{j=0}^{\breve{n}_q-1} \left( \frac{\pi_j^q}{\breve{\beta}_q^0 + \breve{n}_q - j - 1} - \frac{1}{\breve{\alpha}_q^0 + \breve{\beta}_q^0 + j} \right) - \sum_{j=0}^{\breve{n}_q-1} \left( \frac{\pi_j^q}{(\breve{\beta}_q^0 + \breve{n}_q - j - 1)^2} - \frac{1}{(\breve{\alpha}_q^0 + \breve{\beta}_q^0 + j)^2} \right) \breve{\beta}_q^0 \right),$$

$$l, s = m+1, \ldots, 2m,$$

$$H_{sl} = H_{ls} = \sum_{q=1}^{d} \vartheta_{ql}\vartheta_{qs}\breve{\alpha}_q^0\breve{\beta}_q^0 \sum_{j=0}^{\breve{n}_q-1} \frac{1}{(\breve{\alpha}_q^0 + \breve{\beta}_q^0 + j)^2}, \quad l = 1, \ldots, m, \quad s = m+1, \ldots, 2m,$$

$$G_{l1} = \sum_{q=1}^{d} \vartheta_{ql}\breve{\alpha}_q^0 \sum_{j=0}^{\breve{n}_q-1} \frac{\breve{n}_q - j}{\breve{\alpha}_q^0 + j} \breve{P}_j^q, \quad G_{l2} = -\sum_{q=1}^{d} \vartheta_{ql}\breve{\alpha}_q^0 \sum_{j=0}^{\breve{n}_q-1} \frac{j+1}{\breve{\alpha}_q^0 + j} \breve{P}_{j+1}^q, \quad l = 1, \ldots, m,$$

$$G_{l1} = -\sum_{q=1}^{d} \vartheta_{ql}\breve{\beta}_q^0 \sum_{j=0}^{\breve{n}_q-1} \frac{\breve{n}_q - j}{\breve{\beta}_q^0 + \breve{n}_q - j - 1} \breve{P}_j^q, \quad G_{l2} = \sum_{q=1}^{d} \vartheta_{ql}\breve{\beta}_q^0 \sum_{j=0}^{\breve{n}_q-1} \frac{j+1}{\breve{\beta}_q^0 + \breve{n}_q - j - 1} \breve{P}_{j+1}^q,$$

$$l = m+1, \ldots, 2m,$$

$$\pi_j^q = \sum_{z=0}^{j} \breve{P}_z^q(a^0, b^0), \quad \breve{P}_j^q(a^0, b^0) = C_{\breve{n}_q}^j \frac{B(\breve{\alpha}_q^0 + j, \breve{\beta}_q^0 + \breve{n}_q - j)}{B(\breve{\alpha}_q^0, \breve{\beta}_q^0)}, \quad \breve{\alpha}_q^0 = e^{a^{0T}\vartheta_q}, \quad \breve{\beta}_q^0 = e^{b^{0T}\vartheta_q}.$$

*Proof of Theorem 6.* The log-likelihood function for the BLM is expressed as (Slaton et al. 2000)

$$l(a,b) = \sum_{i=1}^{k} \left( \ln(C_{n_i}^{x_i}) + \sum_{j=0}^{x_i-1} \ln(\alpha_i(a) + j) + \sum_{j=0}^{n_i-x_i-1} \ln(\beta_i(b) + j) - \sum_{j=0}^{n_i-1} \ln(\alpha_i(a) + \beta_i(b) + j) \right). \tag{33}$$

Under the theorem assumptions, the function $l(a, b)$ can be rewritten as

$$l(a,b) = \sum_{q=1}^{d}\sum_{t=1}^{k_q}\left(\ln(C_{\check{n}_q}^{y_t^q}) + \sum_{j=0}^{y_t^q-1}\ln(\check{\alpha}_q(a)+j) + \sum_{j=0}^{\check{n}_q-y_t^q-1}\ln(\check{\beta}_q(b)+j) - \sum_{j=0}^{\check{n}_q-1}\ln(\hat{\alpha}_q(a)+\check{\beta}_q(b)+j)\right),$$

where $k_q$ is a number of clusters with factors vector $\vartheta_q$, $k = \sum_{q=1}^{d} k_q$, $y_t^q$ is the observed number of successes for the cluster type $t$, $X = \bigcup_{q=1}^{d}\{y_1^q, y_2^q, \ldots, y_{k_q}^q\}$, and $\check{\alpha}_q(a) = e^{a^T\vartheta_q}$, $\check{\beta}_q(a) = e^{b^T\vartheta_q}$. Then, transforming the sum by $t$ using approach of Johnson *et al.* (1996) for the BBM likelihood system derivation yields

$$l(a,b) = \lambda + \sum_{q=1}^{d} k_q \sum_{j=0}^{\check{n}_q-1}\Big((1-F_j^q)\cdot\ln(\check{\alpha}_q(a)+j) + F_j^q\cdot\ln(\check{\beta}(b)+\check{n}_q-j-1) - \ln(\check{\alpha}_q(a)+\check{\beta}_q(b)+j)\Big),$$

where $\lambda$ is some constant, $F_j^q = \sum_{z=0}^{j} f_z^q$, and $f_z^q$ is a relative frequency of the value $z$ occurrence in a sample $\{y_1^q, y_2^q, \ldots, y_{k_q}^q\}$. Let us use the following asymptotic property of $f_z^q$ (Ivchenko and Medvedev 1984): $f_z^q = \tilde{P}_z^q + O_P(1/\sqrt{k_q})$, where $\tilde{P}_z^q$ is the corresponding theoretical probability. Then, using the properties of $O_P(.)$ and the assumption that the factors $\{\vartheta_1, \vartheta_2, \ldots, \vartheta_d\}$ are equiprobable, it can be proved that for $k \to \infty$, the ML-estimator maximizes the following function

$$l_1(a,b) = \sum_{q=1}^{d}\sum_{j=0}^{\check{n}_q-1}\Big((1-\tilde{\pi}_j^q)\cdot\ln(\check{\alpha}_q(a)+j) + \tilde{\pi}_j^q\cdot\ln(\check{\beta}_q(b)+\check{n}_q-j-1) - \ln(\check{\alpha}_q(a)+\check{\beta}_q(b)+j)\Big) + O_P\left(\frac{1}{\sqrt{k}}\right),$$

where $\tilde{\pi}_j^q = \sum_{z=0}^{j} \tilde{P}_z^q(a^0, b^0, \varepsilon_0, \varepsilon_1)$, and $\tilde{P}_z^q(a^0, b^0, \varepsilon_0, \varepsilon_1)$ are the elements of the probability row for the DBBD with the parameters $\check{n}_q, \check{\alpha}_q, \check{\beta}_q, \varepsilon_0, \varepsilon_1$ (see Theorem 1):

$$\tilde{P}_z^q(a^0, b^0, \varepsilon_0, \varepsilon_1) = \sum_{l=1}^{\check{n}_q} w_{zl}^q(\varepsilon_0, \varepsilon_1)\cdot\check{P}_z^q(a^0, b^0), \quad w_{zj}^q = \sum_{l=\max(z,j)}^{\min(\check{n}_q, z+j)} C_z^{l-z} C_{\check{n}_q-j}^{l-j}\varepsilon_0^{l-j}(1-\varepsilon_0)^{\check{n}_q-l}\varepsilon_1^{l-z}(1-\varepsilon_1)^{j+z-l}.$$

Besides, it can be proved that the following asymptotic expansions for $\tilde{P}_z^q$ hold

$$\tilde{P}_z^q = \check{P}_z^q + \Big((\check{n}_q - z+1)\check{P}_{z-1}^q - (\check{n}_q - z)\check{P}_z^q\Big)\varepsilon_0 + \Big((z+1)\check{P}_{z+1}^q - z\check{P}_z^q\Big)\varepsilon_1 + o(\varepsilon_0) + o(\varepsilon_1), \quad z = 0, 1, \ldots, \check{n}_q, \quad (34)$$

where $\check{P}_{-1}^q = \check{P}_{\check{n}_q+1}^q = 0$. Since the ML-estimator is a solution of the optimization problem $l_1(a, b) \to \max$, then the corresponding partial derivable are equal to zero:

$$\sum_{q=1}^{d}\vartheta_q\check{\alpha}_q(a)\sum_{j=0}^{\check{n}_q-1}\left(\frac{1-\tilde{\pi}_j^q(a,b,\varepsilon_0,\varepsilon_1)}{\check{\alpha}_q(a)+j} - \frac{1}{\check{\alpha}_q(a)+\check{\beta}_q(b)+j}\right) + \mathbf{1}_m\cdot O_P\left(\frac{1}{\sqrt{k}}\right) = \mathbf{0}_m, \quad (35)$$

$$\sum_{q=1}^{d} \vartheta_q \breve{\beta}_q(a) \sum_{j=0}^{\breve{n}_q-1} \left( \frac{\tilde{\pi}_j^q(a,b,\varepsilon_0,\varepsilon_1)}{\breve{\beta}_q(a)+\breve{n}_q-j-1} - \frac{1}{\breve{\alpha}_q(a)+\breve{\beta}_q(b)+j} \right) + \mathbf{1}_m \cdot O_P\left(\frac{1}{\sqrt{k}}\right) = \mathbf{0}_m, \quad (36)$$

where $\mathbf{0}_m$ is a vector of zeros of size $m$. Linearizing this system w.r.t. $\Delta a(\varepsilon_0,\varepsilon_1)$, $\Delta b(\varepsilon_0,\varepsilon_1)$ and expressing the biases from the linearized system concludes the proof.

□

**Expressions for Theorem 7.** In the theorem statement, the following notation is used:

$$J = \begin{pmatrix} J^{Aa} & J^{Ab} \\ J^{Ba} & J^{Bb} \end{pmatrix}, \quad (g_\varepsilon)^T = (g^a, g^b)^T,$$

where

$$J_{ls}^{Aa} = \sum_{i=1}^{k} Z_{il} Z_{is} \alpha_i^0 \left( \sum_{j=0}^{x_i-1} \frac{1}{\alpha_i^0+j} - \sum_{j=0}^{n_i-1} \frac{1}{\alpha_i^0+\beta_i^0+j} - \alpha_i^0 \left( \sum_{j=0}^{x_i-1} \frac{1}{(\alpha_i^0+j)^2} - \sum_{j=0}^{n_i-1} \frac{1}{(\alpha_i^0+\beta_i^0+j)^2} \right) \right),$$

$$J_{ls}^{Bb} = \sum_{i=1}^{k} Z_{il} Z_{is} \beta_i^0 \left( \sum_{j=0}^{n_i-x_i-1} \frac{1}{\beta_i^0+j} - \sum_{j=0}^{n_i-1} \frac{1}{\alpha_i^0+\beta_i^0+j} - \beta_i^0 \left( \sum_{j=0}^{n_i-x_i-1} \frac{1}{(\beta_i^0+j)^2} - \sum_{j=0}^{n_i-1} \frac{1}{(\alpha_i^0+\beta_i^0+j)^2} \right) \right),$$

$$J_{ls}^{Ab} = J_{ls}^{Ba} = \sum_{i=1}^{k} Z_{il} Z_{is} \alpha_i^0 \beta_i^0 \sum_{j=0}^{n_i-1} \frac{1}{(\alpha_i^0+\beta_i^0+j)^2}, \qquad l,s = 1,2,\ldots,m.$$

$$g_l^a = \sum_{i=1}^{k} Z_{il} \alpha_i^0 \left( \frac{-x_i(\beta_i^0+n_i-x_i)}{(\alpha_i^0+x_i-1)^2} \varepsilon_0 + \frac{n_i-x_i}{\beta_i^0+n_i-x_i-1} \varepsilon_1 \right),$$

$$g_l^b = \sum_{i=1}^{k} Z_{il} \beta_i^0 \left( \frac{x_i}{\alpha_i^0+x_i-1} \varepsilon_0 + \frac{(n_i-x_i)(\alpha_i^0+x_i)}{(\beta_i^0+n_i-x_i-1)^2} \varepsilon_1 \right).$$

*Proof of Theorem 7.* Using the asymptotic expansion (34) and the properties of the BBD (Johnson *et al.* 1996), the log-likelihood function $l_\varepsilon(a,b,X,\varepsilon_0,\varepsilon_1)$ can be expressed in the following asymptotic form

$$l_\varepsilon(a,b,X,\varepsilon_0,\varepsilon_1) = l(a,b,X) + e(a,b,X,\varepsilon_0,\varepsilon_1) + o(\varepsilon_0) + o(\varepsilon_1),$$

where

$$e(a,b,X,\varepsilon_0,\varepsilon_1) = \sum_{i=1}^{k} \left( \left( x_i \frac{\beta_i(b)+n_i-x_i}{\alpha_i(a)+x_i-1} - (n_i-x_i) \right) \cdot \varepsilon_0 + \left( (n_i-x_i) \frac{\alpha_i(a)+x_i}{\beta_i(b)+n_i-x_i-1} - x_i \right) \cdot \varepsilon_1 \right),$$

and $l(.)$ is defined by (33). Let us note that, when ignoring the distortions, the ML-estimator $\tilde{a}(X,\varepsilon_0,\varepsilon_1)$, $\tilde{b}(X,\varepsilon_0,\varepsilon_1)$ is the solution of the optimization problem $l(a,b) \to$ max. However, when taking the distortions into account, the ML-estimator $\tilde{a}^0(X,\varepsilon_0,\varepsilon_1)$, $\tilde{b}^0(X,\varepsilon_0,\varepsilon_1)$ is the solution of another problem: $l_\varepsilon(a,b) \to$ max. Let us denote

$$y = \begin{pmatrix} a \\ b \end{pmatrix}, \quad \tilde{y}(X,\varepsilon_0,\varepsilon_1) = \begin{pmatrix} \tilde{a}(X,\varepsilon_0,\varepsilon_1) \\ \tilde{b}(X,\varepsilon_0,\varepsilon_1) \end{pmatrix}, \quad \tilde{y}^0(X,\varepsilon_0,\varepsilon_1) = \begin{pmatrix} \tilde{a}^0(X,\varepsilon_0,\varepsilon_1) \\ \tilde{b}^0(X,\varepsilon_0,\varepsilon_1) \end{pmatrix}, \quad y^0 = \begin{pmatrix} a^0 \\ b^0 \end{pmatrix}.$$

It can be proved that, in the neighborhood of $\tilde{y}^0$, $\frac{\partial l}{\partial y}(\tilde{y}^0) + J(\tilde{y}^0)\cdot(\tilde{y}-\tilde{y}^0) + o(\tilde{y}-\tilde{y}^0) = \mathbf{0}_{2m}$. On the other hand, $\frac{\partial l}{\partial y}(\tilde{y}^0) = -g_e(\tilde{y}^0) + \mathbf{1}_{2m}(o(\varepsilon_0) + o(\varepsilon_1))$, where

$$\frac{\partial l(a,b,X)}{\partial a} = \sum_{i=1}^{k} Z_i \alpha_i(a) \left( \sum_{j=0}^{x_i-1} \frac{1}{\alpha_i(a)+j} - \sum_{j=0}^{n_i-1} \frac{1}{\alpha_i(a)+\beta_i(a)+j} \right), \tag{37}$$

$$\frac{\partial l(a,b,X)}{\partial b} = \sum_{i=1}^{k} Z_i \beta_i(a) \left( \sum_{j=0}^{n_i-x_i-1} \frac{1}{\beta_i(a)+j} - \sum_{j=0}^{n_i-1} \frac{1}{\alpha_i(a)+\beta_i(a)+j} \right). \tag{38}$$

Then, using the above expressions and the asymptotic property of the ML-estimator $\tilde{a}^0 = a^0 + \mathbf{1}_m \cdot O_P(1/\sqrt{k})$, $\tilde{b}^0 = b^0 + \mathbf{1}_m \cdot O_P(1/\sqrt{k})$ completes the proof. □

**MLS-estimation of BLM parameters.** The partial derivatives of the log-function $l(a,b,X,\varepsilon_0,\varepsilon_1)$ are computed as

$$\frac{\partial l}{\partial a_r} = \sum_{i=1}^{k}\sum_{j=0}^{n_i} \frac{w^i_{x_i j}(\varepsilon_0,\varepsilon_1)\cdot \partial P^i_j(a,b)/\partial a_r}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)}, \quad \frac{\partial l}{\partial b_r} = \sum_{i=1}^{k}\sum_{j=0}^{n_i} \frac{w^i_{x_i j}(\varepsilon_0,\varepsilon_1)\cdot \partial P^i_j(a,b)/\partial b_r}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)},$$

$$\frac{\partial l}{\partial \varepsilon_0} = \sum_{i=1}^{k}\sum_{j=0}^{n_i} \frac{\partial w^i_{x_i j}(\varepsilon_0,\varepsilon_1)/\partial \varepsilon_0 \cdot P^i_j(a,b)}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)}, \quad \frac{\partial l}{\partial \varepsilon_1} = \sum_{i=1}^{k}\sum_{j=0}^{n_i} \frac{\partial w^i_{x_i j}(\varepsilon_0,\varepsilon_1)/\partial \varepsilon_0 \cdot P^i_j(a,b)}{\sum_{t=0}^{n_i} w^i_{x_i t}(\varepsilon_0,\varepsilon_1)\cdot P^i_t(a,b)},$$

where

$$\frac{\partial P^i_j(a,b)}{\partial a_r} = P^i_j(a,b) Z_{ir}\, \alpha_i(a) \left( \sum_{l=0}^{j-1} \frac{1}{\alpha_i(a)+l} - \sum_{l=0}^{n_i-1} \frac{1}{\alpha_i(a)+\beta_i(a)+l} \right),$$

$$\frac{\partial P^i_j(a,b)}{\partial b_r} = P^i_j(a,b) Z_{ir}\, \beta_i(a) \left( \sum_{l=0}^{n_i-j-1} \frac{1}{\beta_i(a)+l} - \sum_{l=0}^{n_i-1} \frac{1}{\alpha_i(a)+\beta_i(a)+l} \right),$$

and $\partial w^i/\partial \varepsilon_0$, $\partial w^i/\partial \varepsilon_1$ are defined above (see the MLS-estimator for the BBM).

*Proof of Theorem 8.* Under the distortions, the mean square error of forecasting for the classical Bayes predictor can be expressed as

$$\tilde{r}_i^2 = \mathbf{E}\{(p_i - (\alpha_i^0 + x)/(\alpha_i^0 + \beta_i^0 + n_i))^2\},$$

where $p_i$ is the beta random variable with the parameters $\alpha_i^0, \beta_i^0$, and the variable $x$ (the distorted sum of binary responses) follows the DBBD with the parameters $n_i, \alpha_i^0, \beta_i^0, \varepsilon_0, \varepsilon_1$. Simplifying the latter expression leads to

$$\tilde{r}_i^2 = \mathbf{E}\{p_i^2\} - 2\frac{\alpha_i^0\mathbf{E}\{p_i\} + \mathbf{E}\{xp_i\}}{\alpha_i^0 + \beta_i^0 + n_i} + \frac{\alpha_i^{0\,2} + 2\alpha_i^0\mathbf{E}\{x\} + \mathbf{E}\{x^2\}}{(\alpha_i^0 + \beta_i^0 + n_i)^2}, \tag{39}$$

where $\mathbf{E}\{p_i\} = \alpha_i^0/(\alpha_i^0 + \beta_i^0)$, $\mathbf{E}\{p_i^2\} = \alpha_i^0(\alpha_i^0 + 1)/((\alpha_i^0 + \beta_i^0)(\alpha_i^0 + \beta_i^0 + 1))$, and the mathematical expectations of the random variables $x$, $xp_i$, $x^2$ are

$$\mathbf{E}\{x\} = n\varepsilon_0 + n(1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i\}, \qquad \mathbf{E}\{xp_i\} = n\varepsilon_0\cdot\mathbf{E}\{p_i\} + n(1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i^2\}, \tag{40}$$

$$\mathbf{E}\{x^2\} = \mathbf{E}\{x\} + n(n-1)\left(\varepsilon_0^2 + 2\varepsilon_0(1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i\} + (1-\varepsilon_0-\varepsilon_1)\cdot\mathbf{E}\{p_i^2\}\right). \tag{41}$$

Substituting these formulas to (39) and simplifying the corresponding expression proves the theorem. □

*Proof of Theorem 9.* Following the proof of Theorem 8, the mean square error of forecasting for the classical Bayes predictor under the distortions (when using the estimates $\hat{\alpha}_i, \hat{\beta}_i$) can be expressed as

$$\tilde{r}_i^2 = \mathbf{E}\{p_i^2\} - 2\frac{\hat{\alpha}_i\mathbf{E}\{p_i\} + \mathbf{E}\{xp_i\}}{\hat{\alpha}_i + \hat{\beta}_i + n_i} + \frac{\hat{\alpha}_i^2 + 2\hat{\alpha}_i\mathbf{E}\{x\} + \mathbf{E}\{x^2\}}{(\hat{\alpha}_i + \hat{\beta}_i + n_i)^2}, \tag{42}$$

where the mathematical expectations $\mathbf{E}\{x\}, \mathbf{E}\{xp_i\}, \mathbf{E}\{x^2\}$ are defined by expressions (40), (41). Then, collecting the coefficients of $\varepsilon_0, \varepsilon_1$ and $\varepsilon_0^2$, $\varepsilon_0\varepsilon_1$, $\varepsilon_1^2$ in expression (42) taking into account the notation (22), (23) proves the theorem. □

*Proof of Theorem 10.* Using the Bayes formula and Theorem 1, the posterior p.d.f. of the random variable $p_i$ is expressed as:

$$f_{p_i}(x|s,\varepsilon_0,\varepsilon_1) = \frac{\sum_{r=0}^{n_i} w_{sr}^i(\varepsilon_0,\varepsilon_1)\cdot C_{n_i}^r x^r(1-x)^{(n_i-r)}\cdot B(\alpha_i^0,\beta_i^0)^{-1}x^{\alpha_i^0-1}(1-x)^{\beta_i^0-1}}{\int_0^1\sum_{r=0}^{n_i} w_{sr}^i(\varepsilon_0,\varepsilon_1)\cdot C_{n_i}^r y^r(1-x)^{(n_i-r)}\cdot B(\alpha_i^0,\beta_i^0)^{-1}y^{\alpha_i^0-1}(1-y)^{\beta_i^0-1}dy}.$$

Simplifying this formula using the properties of the beta distribution (Johnson *et al.* 1996) leads to the expression for the forecast p.d.f. (27). Then, calculating the mean of this distribution taking into account the properties of the DBBD gives the predictor

(24). The mean square error of forecasting (26) is derived using the technique given in the proof of Theorem 8 for the obtained predictor. □

## 8 References


Agresti, A., Booth, J. G., Hobert, J. P. and Caffo, B. (2000). Random effects modeling of categorical response data. *Sociological Methodology*, 30, 27-80.

Brooks, S. P. (2001). On Bayesian analysis and finite mixtures for proportions. *Statistics and Computing*, 11, 179-190.

Collet, D. (2002). *Modeling Binary Data*. London: Champton and Hall/CRC.

Coull, B. A. and Agresti, A. (2000). Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, 56, 73-80.

Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of Royal Statistical Society*, 50B, 225-265.

Danaher, P. J. (1992). Some statistical modeling problems in the advertising industry: a look at media exposure distributions. *The American Statistician*, 46, 241-253.

Demidovich B. P. and Maron, I. A. (1970). *Basics of Computational Mathematics*. Moscow (in Russian).

Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford : University Press.

Gaba, A. and Winkler, R. L. (1992). Implication of errors in survey data: a Bayesian model. *Management Science.*, 38 (7), 913-925.

Gill, P. S. (2001). A robust mixed linear model analysis for longitudinal data. *Statistics in Medicine*, 19, 975-987.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistics*. New York: John Wiley and Sons.

Heckman, J. J. and Willis, R. J. (1977). A beta logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy*, 85 (11), 27-58.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Ivchenko, G. I. and Medvedev, U. I. (1984). *Mathematical Statistics*. Moscow (in Russian).

Johnson, N. L., Kotz, S. and Kemp, A. W. (1996). *Univariate Discrete Distributions*. Wiley-Interscience, New York.

Kharin, Yu. (1996). *Robustness in Statistical Pattern Recognition*. Kluwer Academic Publishers, Dordrecht.

Kordzakhia, N., Mishra, G. D. and Reiersolmoen, L. (2001). Robust estimation in the logistic regression model. *Journal of Statistical Planning and Inference*, 98, 211-223.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Nathan, D. (1999). A beta-logistic model of presidential influence on voting on civil rights issues in the house of representatives, 1960-1988. *Midwest Annual Meeting Methodology Panel Working Papers*, 32-37.

Neuhaus, J. M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86, 843-855.

Neuhaus, J. M. (2002). Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics*, 58, 675-683.

Pearson, E. S. (1925). Bayes' theorem in the light of experimental sampling. *Biometrika*, 17, 338-442.

Pfeifer, P. E. (1998). On using the beta-logistic model to update response probabilities given nonresponse. *Journal of Interactive Marketing*, 12 (2), 23-32.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1033-1048.

Prentice, R. L. (1986). Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors. *Journal of the American Statistical Association*, 81, 321-327.

Ruckstuhl, A. F. and Welsh, A. H. (2001). Robust fitting of the binomial model. *The Annals of Statistics*, 29, 1117-1136.

Sissors, J. and Lincoln, B. (1994). *Advertising Media Planning*. NTC Business Books. Lincolnwood.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as a variable between the sets of trials. *Journal of the Royal Statistical Society*, 10B, 257-261.

Slaton, T. L., Piegorsch, W. W. and Durham, S. D. (2000). Estimation and testing with overdispersed proportions using the beta-logistic regression model of Heckman and Willis. *Biometrics*, 56(1), 125-133.

Swartz, T., Haitovsku, Y., Vexler, A. and Yang, T. Bayesian identifiability and misclassification in multinomial data. *The Canadian Journal of Statistics*, 32, 2004, to appear.

Tripathi, R. C., Gupta, R. C. and Gurland, J. (1994). Estimation of parameters in the beta binomial model. *Ann. Inst. Statist. Math.*, 46, 317-331.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Model. *Econometrica*, 50(1), 1-26.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79-86.