On Regression Models for Discrete Difference Distributions

Ghadah Alomani [a]*, Abdulhamid A. Alzaid [b] and Maha A. Omair [b]

[a] *Department of Mathematical Sciences, College of Sciences, Princess Nourah bint Abdulrahman University, Kingdom of Saudi Arabia*
[b] *Department of Statistics and Operations Research, College of Sciences, King Saud University, Kingdom of Saudi Arabia*
** Corresponding author. E-mail address: gaalomani@pnu.edu.sa.*

ABSTRACT

We proposed five different regression models for discrete dependent variables that take both positive and negative values using the Skellam, skew Laplace, trinomial difference, and extended binomial distributions. We applied these models to data from England Premier League and compared their performances.

Keywords: Generalized Linear Model; Regression model; Skellam distribution; Skew Laplace distribution; Trinomial Difference distribution; Extended Binomial distribution.

1. INTRODUCTION

Count regression models as Binary, Poisson and negative binomial regression models are widely used to model count variables that take values in the set of nonnegative integers. Nelder and Wedderburn (1972) showed that these models form a subset of the general class of generalized linear models (GLM). The book of McCullagh and Nelder (1989) is one of the standard books for these models. Hilbe (2011) presented the negative binomial model in detail. Extension of these models to vector of parameters set up can be found in the book of Yee (2015). Nowadays, most packages provide programs that analyze data using GLM. For these models, it is assumed that a function of the mean is linked linearly to the set of explanatory variables and the distribution of the dependent variable is a member of the exponential family distributions. Recently, the difference between two nonnegative integer random variables has attracted the attention of many researchers. The most popular distribution on Z is the Skellam distribution defined by taking the difference of two Poisson random variables. Karlis and Ntzoufras (2006) used the Skellam and the zero inflated Skellam distributions to model an application from dental epidemiology. Karlis and Ntzoufras (2009) also used the Skellam distribution to model the difference of the number of goals in football games. Alzaid and Omair (2010) have used the Skellam distribution for Stock market and hospital occupancy in a nursery intensive care unit. Inusah and Kozubowski (2006) studied discrete skew Laplace distribution (DL). Ong ***et al.*** (2008) defined the difference between two discrete random variables from the Panjer family. Alzaid and Omair (2012) introduced the extended binomial distribution as an extension of the binomial distribution to allow for negative values. Omair ***et al.*** (2016) defined the trinomial difference distribution and used motorcycle accident data for fitting this distribution. Very limited research has considered the regression models using distribution on Z. Karlis and Ntzoufras (2009) applied the Bayesian methodology for the Skellam's distribution for the goal difference using covariates.

This paper is concerned with developing regression models for such distributions. Our approach mimics that of the GLM. The paper is organized as follows: In the rest of this section the basic assumptions of the GLM are given. We developed models for Skellam, skew Laplace, trinomial difference and extended binomial distributions in Section 2. In Section 3, we applied models of Section 2 to a set of data drawn from the English Premier League.

The GLM mainly model one parameter (the mean) of the underlying distribution which is a member of the exponential family. The proposed models in this paper provide greater flexibility.

Nelder and Wederburn (1972) considered a generalized linear model (GLM) in a unified way.

Let $Y_1, Y_2, \ldots, Y_n$ be a set of independent random variables. The basic assumptions of GLM are as follows:

1) The distribution of each Y_i has the conical exponential family form which depends on parameter θ_i (where θ_i is different) with:

$$f(y_i, \theta_i) = exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \tag{1}$$

where b, c and d are known functions.

2) All Y_i 's have the same distributions and the joint probability density function of $Y_1, Y_2, \ldots, Y_n$ is

$$\begin{aligned} f(y_1, \ldots, y_n; \theta_1, \ldots, \theta_n) &= \prod_{i=1}^{n} exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \\ &= exp\Big[\sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)\Big]. \end{aligned} \tag{2}$$

3) If $\mu_i = E(Y_i)$ which is the function of θ_i then it is assumed that there is a monotonic differential function say g such that

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \tag{3}$$

g is named the link function where the vector $\mathbf{x}_i^T = (1\, x_{i1} \ldots x_{ip-1})$ be a $1 \times p$ vector of explanatory variables corresponding to the observation i and $\boldsymbol{\beta} = (\beta_0 \ldots \beta_{p-1})^T$ is $p \times 1$ vector of the parameters. The vector $\mathbf{x}_i^T$ is the ith row of the design matrix $\mathbf{X}$ which is defined as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1p-1} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \ldots & x_{np-1} \end{bmatrix}$$

2. REGRESSION MODEL FOR DISTRIBUTIONS ON Z

Unlike the count model on the set of non-negative integers, the mean of count difference variables can take positive and negative values and hence can also be modeled linearly using identity link. Let $z_1, z_2, \ldots, z_n$ be a set of random variables and let $\mathbf{x}_i^T = (1\, x_{i1} \ldots x_{ip-1})$ be a $1 \times p$ vector of explanatory variables.

2.1 Skellam Model

Unlike the Poisson distribution, the Skellam distribution has two parameters. These two parameters are linked to the mean and the variance as in the following equations. Let z~Skellam (θ_1, θ_2) then

$$P(Z=z) = e^{-\theta_1-\theta_2}\left(\frac{\theta_1}{\theta_2}\right)^{\frac{z}{2}} I_z\left(2\sqrt{\theta_1\theta_2}\right), \quad z = 0, \pm1, \pm2, \ldots \tag{4}$$

where $I_z(x)$ is the modified Bessel function. The mean and the variance are $\mu = \theta_1 - \theta_2$ and $\sigma^2 = \theta_1 + \theta_2$.

$$\mathrm{P(Z=z)} = \exp(-\sigma^2)\left(\frac{\sigma^2+\mu}{\sigma^2-\mu}\right)^{\frac{z}{2}} \mathrm{I_z}\left(\sqrt{\sigma^4-\mu^2}\right), \text{z=0}, \pm1, \pm2, \ldots \tag{5}$$

Therefore, we can either model the mean when the variance is not affected by the explanatory variables or model the two parameters to catch the effect of the explanatory variables on the variance.

1) Modeling the Mean

To model the mean, it is better to use the probability function (5). As in the GLM, we assume there is a link function that connects the mean with the parameter as in

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i\,. \tag{6}$$

The likelihood function of this model is

$$L(\mathbf{z}, \mu, \sigma^2) = \prod_{i=1}^{n} \exp(-\sigma^2)\left(\frac{\sigma^2+\mu_i}{\sigma^2-\mu_i}\right)^{\frac{z_i}{2}} I_{z_i}\left(\sqrt{\sigma^4-\mu_i^2}\right), \quad z_i = 0, \pm1, \pm2, \ldots, i = 1, 2, \ldots, n.$$

The score functions are given by

$$U_k = \frac{\partial logL}{\partial \boldsymbol{\beta}_k} = \frac{\partial logL}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}_k} \quad i = 1, 2, .., n, \quad k = 0, 1, \ldots, p-1 \tag{7}$$

To find the estimate of the regression parameters $\boldsymbol{\beta}$ and σ, we solve the equations

$$U_k = \frac{\partial l}{\partial \boldsymbol{\beta}_k} = \sum_{i=1}^{n}\Big(\frac{z_i}{\hat{\sigma}^2+\hat{\mu}_i} - \frac{\hat{\mu}_i}{\sqrt{\hat{\sigma}^4-\hat{\mu}_i^2}} R_{z_i}(\sqrt{\hat{\sigma}^4-\hat{\mu}_i^2})\Big)\mathbf{x}_i^T = 0 \quad (8)$$

$$\frac{\partial l}{\partial \sigma^2} = \sum_{i=1}^{n}\Big(-1 + \frac{z_i}{\hat{\sigma}^2+\hat{\mu}_i} + \frac{\hat{\sigma}^2}{\sqrt{\hat{\sigma}^4-\hat{\mu}_i^2}} R_{z_i}(\sqrt{\hat{\sigma}^4-\hat{\mu}_i^2})\Big) = 0, \quad (9)$$

where $R_z(x) = \dfrac{I_{z+1}(x)}{I_z(x)}$.

For this model, two link functions are of interest, namely the identity

$$g_1(\mu_i) = \mu_i, \quad (10)$$

and the log function

$$g_2(\mu_i) = log\, \mu_i. \quad (11)$$

A possible interpretation of the identity model (10) is as follows

$$\begin{aligned} E(Y_i) &= \mathbf{x}^T\boldsymbol{\beta} \\ Y_i = Z_i &+ \varepsilon_i, \end{aligned}$$

where $Z_i(\mathbf{x}^T\boldsymbol{\beta}) \sim Skellam(\mathbf{x}^T\boldsymbol{\beta}, \sigma_1^2)$ is independent of $\varepsilon_i \sim Skellam(0, \sigma_2^2)$ and $\sigma_1^2 + \sigma_2^2 = \sigma^2$.

2) Modeling the parameters

For this model, we considered the standard Skellam probability mass function (4). We assumed here that the parameters θ_1 and θ_2 are linked with the explanatory variables by

$$\theta_{1i} = \exp(\mathbf{x}_i^T\boldsymbol{\beta}_1) \quad \text{and} \quad \theta_{2i} = \exp(\mathbf{x}_i^T\boldsymbol{\beta}_2)$$

Here, we assumed the independence of the parameters to be

$$\boldsymbol{\beta_1} = \begin{bmatrix} \beta_{10} \\ \vdots \\ \beta_{1p-1} \end{bmatrix} \text{ and } \boldsymbol{\beta_2} = \begin{bmatrix} \beta_{20} \\ \vdots \\ \beta_{2p-1} \end{bmatrix}$$

The corresponding likelihood function for (4) is given by

$$L(\mathbf{z}, \theta_{1i}, \theta_{2i}) = \prod_{i=1}^{n} \exp(-\theta_{1i} - \theta_{2i})\left(\frac{\theta_{1i}}{\theta_{2i}}\right)^{\frac{z_i}{2}} I_{z_i}\left(2\sqrt{\theta_{1i}\theta_{2i}}\right). \quad (12)$$

Hence, the score functions are given by

$$U_{lk} = \frac{\partial l}{\partial \beta_{lk}} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \theta_{ki}} \frac{\partial \theta_{ki}}{\partial \beta_{lk}}, \; i = 1, 2, \ldots, n, \; k = 0, 1, \ldots, p-1, \; l = 1, 2. \quad (13)$$

The maximum likelihood estimator of the regression parameter $\boldsymbol{\beta}_l$ is obtained by solving the following nonlinear equations:

$$\frac{\partial l}{\partial \beta_{1k}} = \sum_{i=1}^{n}\left(-\hat{\theta}_{1i} + z_i + \sqrt{\hat{\theta}_{1i}\hat{\theta}_{2i}}\, R_{z_i}\left(2\sqrt{\hat{\theta}_{1i}\hat{\theta}_{2i}}\right)\right)x_{ik} = 0 \tag{14}$$

$$\frac{\partial l}{\partial \beta_{2k}} = \sum_{i=1}^{n}\left(-\hat{\theta}_{2i} + z_i + \sqrt{\hat{\theta}_{1i}\hat{\theta}_{2i}}\, R_{z_i}\left(2\sqrt{\hat{\theta}_{1i}\hat{\theta}_{2i}}\right)\right)x_{ik} = 0, \tag{15}$$

where $\theta_{li} = e^{\mathbf{x}_i^T \boldsymbol{\beta}_l}$, $l = 1,2$ and $k = 0,1,\dots,p-1$.

Hypothesis test and confidence interval can be based on the fact the maximum likelihood estimator $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2)$ is asymptotically normally distributed $N(\Theta, I^{-1}(\Theta))$ where

$$I(\Theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

is a $2p \times 2p$ matrix, with

$$I_{ll} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_{l0} \partial \beta_{l0}} & \frac{\partial^2 l}{\partial \beta_{l0} \partial \beta_{l1}} & \cdots & \frac{\partial^2 l}{\partial \beta_{l0} \partial \beta_{lp-1}} \\ \vdots & & & \vdots \\ \frac{\partial^2 l}{\partial \beta_{lp-1} \partial \beta_{l0}} & \frac{\partial^2 l}{\partial \beta_{lp-1} \partial \beta_{l1}} & \cdots & \frac{\partial^2 l}{\partial \beta_{lp-1} \partial \beta_{lp-1}} \end{bmatrix}, l = 1,2, \tag{16}$$

and

$$I_{12} = I_{21} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_{10} \partial \beta_{20}} & \frac{\partial^2 l}{\partial \beta_{10} \partial \beta_{21}} & \cdots & \frac{\partial^2 l}{\partial \beta_{10} \partial \beta_{2p-1}} \\ \vdots & & & \vdots \\ \frac{\partial^2 l}{\partial \beta_{1p-1} \partial \beta_{20}} & \frac{\partial^2 l}{\partial \beta_{1p-1} \partial \beta_{21}} & \cdots & \frac{\partial^2 l}{\partial \beta_{1p-1} \partial \beta_{2p-1}} \end{bmatrix}, \tag{17}$$

are $p \times p$ matrices.

For the present model, we have

$$I_{11} = -\frac{\partial^2 l}{\partial \beta_{1k}\beta_{1j}} = -\sum_{i=1}^{n}\Big\{-\theta_{1i} + \sqrt{\theta_{1i}\theta_{2i}}R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}}) + \theta_{1i}\theta_{2i}R_{z_i+1}(\sqrt{\theta_{1i}\theta_{2i}}) \times R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}}) - \theta_{1i}\theta_{2i}\big(R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}})\big)^2\Big\} \times x_{ik}x_{ij}. \quad (18)$$

$$I_{22} = -\frac{\partial^2 l}{\partial \beta_{2k}\beta_{2j}} = -\sum_{i=1}^{n}\Big\{-\theta_{2i} + \sqrt{\theta_{1i}\theta_{2i}}R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}}) + \theta_{1i}\theta_{2i}R_{z_i+1}(\sqrt{\theta_{1i}\theta_{2i}}) \times R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}}) - \theta_{1i}\theta_{2i}\big(R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}})\big)^2\Big\} \times x_{ik}x_{ij}. \quad (19)$$

$$I_{12} = I_{21} = -\frac{\partial^2 l}{\partial \beta_{1k}\beta_{2j}} = -\sum_{i=1}^{n}\Big\{\sqrt{\theta_{1i}\theta_{2i}}R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}}) + \theta_{1i}\theta_{2i}R_{z_i+1}(\sqrt{\theta_{1i}\theta_{2i}}) \times R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}}) - \theta_{1i}\theta_{2i}\big(R_{z_i}(\sqrt{\theta_{1i}\theta_{2i}})\big)^2\Big\} \times x_{ik}x_{ij}, \quad (20)$$

where k, j=0,1,...,p-1.

2.2 ***Skew Laplace Model***

Let z~skew Laplace (p_1, p_2) then the probability mass function is

$$P(Z = z) = \frac{(1-p_1)(1-p_2)}{1-p_1p_2}\begin{cases} p_1^z & z = 0,1,2,3,... \\ p_2^{-z} & z = 0,-1,-2,-3,..., \end{cases} \quad (21)$$

For a skew Laplace model, the ith observation can be written as

$$f(z_i \mid p_{1i}, p_{2i}) = P(Z = z_i) = \frac{(1-p_{1i})(1-p_{2i})}{1-p_{1i}p_{2i}} p_{1i}^{\frac{|z_i|+z_i}{2}} p_{2i}^{\frac{|z_i|-z_i}{2}}, \quad i = 1,2,...,n \quad (22)$$

where $0 < p_{ki} < 1$, for k=1, 2 and i=1,2,...,n, and the following link functions:

$$p_{li} = \frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}_l}}{1+e^{\mathbf{x}_i^T\boldsymbol{\beta}_l}}, \; l = 1,2.$$

The corresponding likelihood function is

$$L(\mathbf{z}, p_{1i}, p_{2i}) = \prod_{i=1}^{n}\frac{(1-p_{1i})(1-p_{2i})}{1-p_{1i}p_{2i}} p_{1i}^{\frac{|z_i|+z_i}{2}} p_{2i}^{\frac{|z_i|-z_i}{2}}. \quad (23)$$

To obtain the maximum likelihood estimator of $\boldsymbol{\beta}_l$ we solve these following equations:

$$\frac{\partial l}{\partial \beta_{1k}} = \sum_{i=1}^{n}\Big(\frac{\hat{p}_{1i}\hat{p}_{2i}(1-\hat{p}_{1i})}{1-\hat{p}_{1i}\hat{p}_{2i}} - \hat{p}_{1i} + \frac{|z_i|+z_i}{2}(1-\hat{p}_{1i})\Big)x_{ik} = 0 \quad (24)$$

$$\frac{\partial l}{\partial \beta_{2k}} = \sum_{i=1}^{n}\Big(\frac{\hat{p}_{1i}\hat{p}_{2i}(1-\hat{p}_{2i})}{1-\hat{p}_{1i}\hat{p}_{2i}} - \hat{p}_{2i} + \frac{|z_i|-z_i}{2}(1-\hat{p}_{2i})\Big)x_{ik} = 0 \quad (25)$$

where k=0,1,...,p-1.

Hence, the information matrix is given by

$$I_{11} = -\frac{\partial^2 l}{\partial \beta_{1k}\beta_{1j}} = -\sum_{i=1}^{n}\Big(\frac{p_{1i}p_{2i}(1-p_{1i})(1-2p_{1i}+p_{1i}^2 p_{2i})}{(1-p_{1i}p_{2i})^2}$$
$$-\Big(1+\frac{|z_i|+z_i}{2}\Big)p_{1i}(1-p_{1i})\Big)x_{ik}x_{ij} \quad (26)$$

$$I_{22} = -\frac{\partial^2 l}{\partial \beta_{2k}\beta_{2j}} = -\sum_{i=1}^{n}\Big(\frac{p_{1i}p_{2i}(1-p_{2i})(1-2p_{2i}+p_{2i}^2 p_{1i})}{(1-p_{1i}p_{2i})^2}$$
$$-\Big(1+\frac{|z_i|-z_i}{2}\Big)p_{2i}(1-p_{2i})\Big)x_{ik}x_{ij} \quad (27)$$

$$I_{12} = I_{21} = -\frac{\partial^2 l}{\partial \beta_{1k}\beta_{2j}} = -\sum_{i=1}^{n}\Big(\frac{p_{1i}p_{2i}(1-p_{1i})(1-p_{2i})}{(1-p_{1i}p_{2i})^2}\Big)x_{ik}x_{ij} \quad (28)$$

2.3 ***Trinomial Difference Model***

Let z~Trinomial difference (n, α,γ) then the probability distribution function is

$$P(Z=z) = \sum_{j=max(0,-z)}^{\left[\frac{n-z}{2}\right]} \binom{n}{z+j,\, j} \alpha^{z+j}\gamma^{j}(1-\alpha-\gamma)^{n-z-2j} \quad z = 0,\pm1,\pm2,\ldots,\pm n \quad (29)$$

where $n \in Z, 0 \le \alpha,\gamma \le 1$ and $\alpha+\gamma<1$. The ith observation is

$$f(z_i \mid \alpha_i,\gamma_i) = P(Z=z_i) = \sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]} \binom{n}{z_i+j,\, j} \alpha_i^{z_i+j}\gamma_i^{j}(1-\alpha_i-\gamma_i)^{n-z_i-2j}, \quad (30)$$

where $z_i = 0,\pm1,\pm2,\ldots,\pm n$ and i=1,2,...,m with the following link functions:

$$\alpha_i = \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta}_1)}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T\boldsymbol{\beta}_2)} \quad \text{and} \quad \gamma_i = \frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta}_2)}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}_1)+\exp(\mathbf{x}_i^T\boldsymbol{\beta}_2)}$$

The likelihood function for this model is given by

$$L(\mathbf{z},\alpha_i,\gamma_i) = \prod_{i=1}^{m}\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]} \binom{n}{z_i+j,\, j} \alpha_i^{z_i+j}\gamma_i^{j}(1-\alpha_i-\gamma_i)^{n-z_i-2j} \quad (31)$$

The maximum likelihood estimator of $\boldsymbol{\beta}_l$ is given by solving these two equations:

$$\frac{\partial l}{\partial \boldsymbol{\beta}_{1k}}=\sum_{i=1}^{m}\frac{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\hat{\alpha}_i^{z_i+j}\hat{\gamma}_i^{j}(1-\hat{\alpha}_i-\hat{\gamma}_i)^{n-z_i-2j}\left[z_i+j-n\hat{\alpha}_i\right]}{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\hat{\alpha}_i^{z_i+j}\hat{\gamma}_i^{j}(1-\hat{\alpha}_i-\hat{\gamma}_i)^{n-z_i-2j}}\times x_{ik}=0 \quad (32)$$

$$\frac{\partial l}{\partial \boldsymbol{\beta}_{2k}}=\sum_{i=1}^{m}\frac{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\hat{\alpha}_i^{z_i+j}\hat{\gamma}_i^{j}(1-\hat{\alpha}_i-\hat{\gamma}_i)^{n-z_i-2j}\left[j-n\hat{\gamma}_i\right]}{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\hat{\alpha}_i^{z_i+j}\hat{\gamma}_i^{j}(1-\hat{\alpha}_i-\hat{\gamma}_i)^{n-z_i-2j}}\times x_{ik}=0 \quad (33)$$

where k=0, 1,..., p-1.

Thus, the information matrix for this model is given by

$$I_{11}=-\frac{\partial^2 l}{\partial \boldsymbol{\beta}_{1k}\partial \boldsymbol{\beta}_{1j}}=\sum_{i=1}^{m}\frac{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}n\alpha_i^{z_i+j+1}\gamma_i^{j}(1-\alpha_i)(1-\alpha_i-\gamma_i)^{n-z_i-2j}}{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\alpha_i^{z_i+j}\gamma_i^{j}(1-\alpha_i-\gamma_i)^{n-z_i-2j}}\times x_{ik}x_{ij}, \quad (34)$$

$$I_{22}=-\frac{\partial^2 l}{\partial \boldsymbol{\beta}_{2k}\partial \boldsymbol{\beta}_{2j}}=\sum_{i=1}^{m}\frac{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}n\alpha_i^{z_i+j}\gamma_i^{j+1}(1-\gamma_i)(1-\alpha_i-\gamma_i)^{n-z_i-2j}}{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\alpha_i^{z_i+j}\gamma_i^{j}(1-\alpha_i-\gamma_i)^{n-z_i-2j}}\times x_{ik}x_{ij}, \quad (35)$$

$$I_{12}=I_{21}=-\frac{\partial^2 l}{\partial \boldsymbol{\beta}_{1k}\partial \boldsymbol{\beta}_{2j}}=-\sum_{i=1}^{m}\frac{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}n\alpha_i^{z_i+j+1}\gamma_i^{j+1}(1-\alpha_i-\gamma_i)^{n-z_i-2j}}{\sum_{j=max(0,-z_i)}^{\left[\frac{n-z_i}{2}\right]}\binom{n}{z_i+j,j}\alpha_i^{z_i+j}\gamma_i^{j}(1-\alpha_i-\gamma_i)^{n-z_i-2j}}\times x_{ik}x_{ij}, \quad (36)$$

k, j=0,1,...,p-1.

*2.4 **Extended Binomial Model***

Let z~extended binomial (w, *p*, θ) then the probability distribution function is

$$P(Z=z)=\frac{p^z\,q^{w-z}\;{}_0\tilde{F}_1(;z+1;p^2\theta)\;{}_0\tilde{F}_1(;w-z;q^2\theta)}{{}_0\tilde{F}_1(;w+1;\theta)} \tag{37}$$

where

where $z=0,\pm1,\pm2,\ldots$.

The ith observation is

$$f(z_i|p_i\,\theta_i)=p(Z=z_i)$$
$$=\frac{p_i^{z_i}\,q_i^{w-z_i}\;{}_0\tilde{F}_1(;z_i+1;p_i^2\theta_i)\;{}_0\tilde{F}_1(;w-z_i;q_i^2\theta_i)}{{}_0\tilde{F}_1(;w+1;\theta_i)} \tag{38}$$

Where ${}_0\tilde{F}_1(;w;\theta)$ is the regularized generalized hypergeometric function and i=1, 2,...,n, with the following link functions:

$$p_i=\frac{\exp(\mathbf{x}_i^T\boldsymbol{\beta}_1)}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta}_1)} \quad \text{and} \quad \theta_i=\exp(\mathbf{x}_i^T\boldsymbol{\beta}_2).$$

The maximum likelihood estimator of $\boldsymbol{\beta}_l$ is given by

$$L(\mathbf{z},p_i,\theta_i)=\prod_{i=1}^{n}\frac{p_i^{z_i}\,q_i^{w-z_i}\;{}_0\tilde{F}_1(;z_i+1;p_i^2\theta_i)\;{}_0\tilde{F}_1(;w-z_i;q_i^2\theta_i)}{{}_0\tilde{F}_1(;w+1;\theta_i)} \tag{39}$$

To find the estimator of β_k we solve these non-linear equations:

$$\frac{\partial l}{\partial\beta_{1k}}=\sum_{i=1}^{n}\Big(z_i(1-\hat{p}_i)-(w-z_i)\hat{p}_i+2\hat{p}_i^2(1-\hat{p}_i)\hat{\theta}_iH_1(z_i+2,\hat{p}_i^2\hat{\theta}_i) \tag{40}$$
$$-2\hat{p}_i(1-\hat{p}_i)^2\hat{\theta}_iH_1(w-z_i+2,(1-\hat{p}_i)^2\hat{\theta}_i)\Big)x_{ik}=0$$

$$\frac{\partial l}{\partial\beta_{2k}}=\sum_{i=1}^{n}\Big(-\hat{\theta}_iH_1(w+2,\hat{\theta}_i)+\hat{p}_i^2\hat{\theta}_iH_1(z_i+2,\hat{p}_i^2\hat{\theta}_i) \tag{41}$$
$$+(1-\hat{p}_i)^2\hat{\theta}_iH_1(w-z_i+2,(1-\hat{p}_i)^2\hat{\theta}_i)\Big)x_{ik}=0.$$

where $H_k(u_1+1,u_2+1,\ldots,u_k+1,\theta)=\frac{{}_0\tilde{F}_k(u_1+1,u_2+1,\ldots,u_k+1;\theta)}{{}_0\tilde{F}_k(u_1,u_2,\ldots,u_k;\theta)}$.

For the present model, we have

$$
\begin{aligned}
I_{11} = -\frac{\partial^2 l}{\partial\beta_{1k}\beta_{1j}} = -\sum_{i=1}^{n}\Big(& 2p_i^2(1-p_i)(2-3p_i)\theta_i H_1(z_i+2,p_i{}^2\theta_i) + 4p_i^4(1-p_i)^2\theta_i^2 \\
&\times\Big(H_1(z_i+3,p_i{}^2\theta_i)H_1(z_i+2,p_i{}^2\theta_i) - (H_1(z_i+2,p_i^2\theta_i))^2\Big) \\
&-2p_i(1-p_i)(1-4p_i+3p_i^2)\theta_i H_1(w-z_i+2,(1-p_i)^2\theta_i) - \\
&wp_i(1-p_i) + 4p_i^2(1-p_i)^4\theta_i^2\Big(H_1(w-z_i+3,(1-p_i)^2\theta_i)\times \\
&H_1(w-z_i+2,(1-p_i)^2\theta_i) - (H_1(w-z_i+2,(1-p_i)^2\theta_i))^2\Big)\Big)x_{ik}x_{ij}
\end{aligned}
\tag{42}
$$

$$
\begin{aligned}
I_{22} = -\frac{\partial^2 l}{\partial\beta_{2k}\partial\beta_{2j}} = -\sum_{i=1}^{n}\Big(& -\theta_i H_1(w+2,\theta_i) - \theta_i^2\Big(H_1(w+3,\theta_i)H_1(w+2,\theta_i) \\
&-\big(H_1(w+2,\theta_i)\big)^2\Big) + p_i^2\theta_i H_1(z_i+2,p_i^2\theta_i) + \theta_i^2 p_i^2 \\
&\times\Big(H_1(z_i+3,p_i^2\theta_i)H_1(z_i+2,p_i^2\theta_i) - \big(H_1(z_i+2,p_i^2\theta_i)\big)^2\Big) \\
&+(1-p_i)^2\theta_i H_1(w-z_i+2,+(1-p_i)^2\theta_i) + (1-p_i)^4 \\
&\times\theta_i^2\Big(H_1(w-z_i+3,+(1-p_i)^2\theta_i)H_1(w-z_i+2,+(1-p_i)^2\theta_i) \\
&-\big(H_1(w-z_i+2,+(1-p_i)^2\theta_i)\big)^2\Big)\Big)x_{ik}x_{ij}
\end{aligned}
\tag{43}
$$

$$
\begin{aligned}
I_{12} = I_{21} = -\frac{\partial^2 l}{\partial\beta_{1k}\partial\beta_{2j}} = -\sum_{i=1}^{n}\Big(& 2p_i^2(1-p_i)\theta_i H_1(z_i+2,p_i{}^2\theta_i) + 2p_i^4(1-p_i)\theta_i^2 \\
&\times\Big(H_1(z_i+3,p_i{}^2\theta_i)H_1(z_i+2,p_i{}^2\theta_i) - (H_1(z_i+2,p_i^2\theta_i))^2\Big) \\
&-2p_i(1-p_i)^2\theta_i H_1(w-z_i+2,(1-p_i)^2\theta_i) - 2p_i(1-p_i)^4\theta_i^2 \\
&\times\Big(H_1(w-z_i+3,(1-p_i)^2\theta_i)H_1(w-z_i+2,(1-p_i)^2\theta_i) \\
&-(H_1(w-z_i+2,(1-p_i)^2\theta_i))^2\Big)\Big)x_{ik}x_{ij}.
\end{aligned}
\tag{44}
$$

The procedure NLMIXED in SAS software was used to fit these models.

3. APPLICATIONS

3.1 ***Application 1***

The data is obtained from the English Premier League for 2013-2014 seasons. Data were downloaded from http://uk.soccerway.com/. We fitted Skellam, skew Laplace, trinomial difference and the extended binomial models for the goal difference in 2013-2014 seasons. We specified the response variable Y as the goal deference (number of goals by the home team - number of goals by the away team). We used the goal difference (X_1), the total points for the home team (X_2) and the total points for the away team (X_3) in the similar game occurred the last year 2012-2013 as explanatory variables. Note that we eliminated the cases where there are no similar games in 2012-2013 and 2013-2014. We fitted all models, and then we eliminated the non-significant variables. Table 1 displays some descriptive statistics for response and explanatory variables.

Table 1. Descriptive Statistics

Variable	Sample size	Mean	Standard deviation	Minimum	Maximum	Skewness
Y	272	0.401	1.977	-5	7	0.24
X_1	272	0.105	1.739	-6	8	0.40
X_2	272	15.878	15.878	39	89	0.68
X_3	272	15.878	15.878	39	89	0.68

Table 2 displays the fitted models for all parameters. Table 3 shows the fitted models using only significant variables. Based on AIC, Table 2 indicates that the trinomial difference model provided a better fit than the other models. From Table 3, we see that the result does not change when we use only the significant variables. Figure 1 illustrates the results of the goal differences for each game in 2013-2014 season along with the fitted models using all predictors. In Figure 2, we kept only significant predictors in each fitted model.

Table 4 presents the percentage of predicting the correct winning team. That is the argument percentage of the signs of the empirical difference and the signs of the fitted models. From Table 4, we can say that the percentage of compatibility in predicting the exact signs of the differences in goals for Laplace, Skellam and trinomial difference is about the same and higher than the corresponding percentage for the extended binomial distribution. Table 5 presents the percentages of predicting the difference in goal with a margin error of one goal for the four models. Again, the Laplace, Skellam and trinomial difference models are almost compatible in predicting the difference in goals with a margin error of one goal.

*3.2 **Application 2 (Arsenal)***

The data is obtained from the English Premier League for 2013-2014 season. Data were downloaded from http://uk.soccerway.com/. We also considered fitting the goal difference for Arsenal, where the response variable is defined as Y= number of goals of Arsenal - number of goals of the other team in game i. The explanatory variables are the home effect (X_1), the goal difference in the previous year 2012-2013 for the same teams (X_2) and the points for the competing team in the game (X_3). Note that we eliminated the cases where there are different teams in 2012-2013 and 2013-2014. We fitted all models and then we eliminated the non-significant variables. Table 6 displays some descriptive statistics for response and explanatory variables. Table 7 displays the fitted models for all parameters. Table 8 displays the fitted models using only significant variables. Based on AIC, Table 7 indicates that the trinomial difference model provided a better fit than the other models. From Table 8, we see that the result does not change when we use the significant variables. Figure 3 displays the goal difference in each game between Arsinal and other teams in 2013-2014 season along with the fitted models using all predictors. In Figure 4, we kept only significant predictors in each fitted model.

As in the model for all teams, we calculated the percentage of predicting the correct winning team and the percentage of predicting the difference in goals with a margin error of one goal, which are exhibited in Tables 9 and 10. From these two tables, we concluded that the percentage of compatibility in predicting the correct winning team for Laplace, Skellam and trinomial difference is about the same and higher than the corresponding percentage for the extended binomial distribution. However, when we use the significant parameters, the percentage of predicting the exact difference in goals for Laplace and Skellam models is about the same and higher than the corresponding percentage for the other two models. Again, the Laplace, Skellam and trinomial difference models are almost compatible in predicting the difference in goals with a margin error of one goal, while when we use the significant parameters, Laplace model has higher corresponding percentage than the other models.

Table 2: Fitted models for goal difference (2012 -2013) in English Premier League.

parameters	Skellam		Laplace		trinomial difference		extended binomial	
	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value
Intercept (β_{10})	-0.1248	0.7629	-0.3543	0.4663	-2.5427	<.0001	-0.1002	0.7087
Goal difference 2012 -2013 (β_{11})	0.0279	0.5678	0.0225	0.7107	0.0305	0.6699	0.6905	<.0001
Point home (β_{12})	0.0170	0.0005	0.0222	0.0005	0.0248	0.0032	-0.0958	<.0001
Point away (β_{13})	-0.0048	0.4079	-0.0013	0.0572	-0.0035	0.6656	0.0797	<.0001
Intercept (β_{20})	-0.4634	0.3475	-0.7680	0.1843	-2.9049	<.0001	0.6022	0.4630
Goal difference 2012 -2013 (β_{21})	-0.0055	0.9308	-0.0344	0.6406	-0.0041	0.9616	-1.0950	<.0001
Point home (β_{22})	-0.0014	0.8425	-0.0089	0.2561	0.0073	0.5024	0.1936	<.0001
Point away (β_{23})	0.0156	0.0096	0.0194	0.0052	0.0170	0.0346	-0.5019	<.0001
AIC	1100.3		1141.7		1099.0		10406	
BIC	1129.1		1170.6		1127.8		10435	
MSE	4.0264		3.2694		4.3194		10.5004	

Table 3: Fitted models using only significant variables in England Premier League.

parameters	Skellam		Laplace		trinomial difference		extended binomial	
	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value
Intercept (β_{10})					-2.4929	<.0001		
Goal difference 2012 -2013 (β_{11})							0.6674	<.0001
Point home (β_{12})	0.0194	<.0001	0.0200	<0.0001	0.0209	<.0001	-0.0673	<.0001
Point away (β_{13})	-0.0088	0.0088	-0.0168	0.0005			0.0697	<.0001
Intercept (β_{20})					-2.7535	<.0001	0.5999	<.0001
Goal difference 2012 -2013 (β_{21})							-1.0551	<.0001
Point home (β_{22})			-0.0170	0.0018			0.1922	<.0001
Point away (β_{23})	0.0077	<.0001	0.0140	0.0025	0.0211	<.0001	-0.6437	<.0001
AIC	1093.7			1136.1	1092.8			13738
BIC	1104.5			1150.5	1107.2			13763
MSE	3.3052			3.2483	3.2491			12.9732

Table 4: The percentage of predicting the correct winning team.

The model	Percentage (all parameters)	Percentage (significant)
Laplace	56.25%	58.09 %
Skellam	55.88%	56.25 %
Trinomial difference	57.35%	57.35%
Extended binomial	45.59 %	45.59%

Table 5: The percentage of predicting the difference in goals with a margin error of one goal.

The model	With a margin of difference goals equal 1(all parameters)	With a margin of difference goals equal 1 (significant)
Laplace	62.5%	62.13 %
Skellam	62.13%	61.76 %
Trinomial difference	60.29%	60.66%
Extended binomial	34.19%	29.41%

Table 6: Descriptive Statistics.

Variable	Sample size	Mean	Standard deviation	Minimum	Maximum	Skewness
Y	32	0.469	0.373	-6	3	-1.39
X_1	32	0.5	0.0898	0	1	0
X_2	32	0.719	0.299	-2	5	0.73
X_3	32	54.38	2.82	39	89	0.9

Table 7: Detail of the fitted model for data in English Premier League (Arsenal).

parameters	Skellam		Laplace		trinomial difference		extended binomial	
	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value
(β_{10})	2.4953	0.0231	2.1937	0.1078	4.6265	0.0070	-4.0016	<.0001
(β_{11})	-0.3002	0.4954	0.0668	0.9095	-2.6260	0.0016	2.9994	<.0001
(β_{12})	0.1291	0.2570	0.1443	0.4190	0.1916	0.3289	0.3999	<.0001
(β_{13})	-0.0407	0.0772	-0.0432	0.1086	-0.0669	0.0249	0.0484	<.0001
(β_{20})	-0.8072	0.5118	-2.0693	0.2743	2.2600	0.1920	10.6015	<.0001
(β_{21})	-2.1540	0.0102	-1.7544	0.0592	-4.2543	0.0002	-7.0085	<.0001
(β_{22})	0.1874	0.4406	0.1794	0.5657	0.2040	0.5244	-0.0000	0.0089
(β_{23})	0.0248	0.1623	0.0384	0.1756	-0.0195	0.4612	-0.5632	<.0001
AIC	130.6		142.2		127.2		1465.3	
BIC	142.3		153.9		138.9		1477.1	
MSE	2.1237		2.3201		2.0567		10.9064	

Table 8: Fitted models using only significant variables (Arsenal).

parameters	Skellam		Laplace		trinomial difference		extended binomial	
	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value
(β_{10})	2.8874	0.0040	2.9908	0.0165	3.2210	0.0026	-4.0016	<.0001
(β_{11})					-1.0621	0.0125	2.9994	<.0001
(β_{12})							0.3999	<.0001
(β_{13})	-0.0486	0.0248	-0.0554	0.0285	-0.0647	0.0031	0.0484	<.0001
(β_{20})							10.6015	<.0001
(β_{21})	-1.877	0.0146	-1.463	0.0645	-2.8855	0.0007	-7.0085	<.0001
(β_{22})	0.0123	0.0018					-0.0000	0.0089
(β_{23})							-0.5632	<.0001
AIC	124.5		136.3		122.6		1465.3	
BIC	130.4		140.7		128.5		1477.1	
MSE	2.3346		2.7563		2.1560		10.9064	

Table 9: The percentage of predicting the correct winning team.

The model	Percentage (all parameters)	Percentage (significant)
Laplace	71.88%	71.88 %
Skellam	71.88%	71.88 %
Trinomial difference	71.88%	59.38%
Extended binomial	59.38%	59.38%

Table 10: The percentage of predicting the difference in goals with a margin error of one goal.

The model	With a margin of difference goals equal 1 (all parameters)	With a margin of difference goals equal 1 (significant)
Laplace	78.13%	78.13 %
Skellam	78.13%	50 %
Trinomial difference	75%	53.13%
Extended binomial	37.5%	37.5%

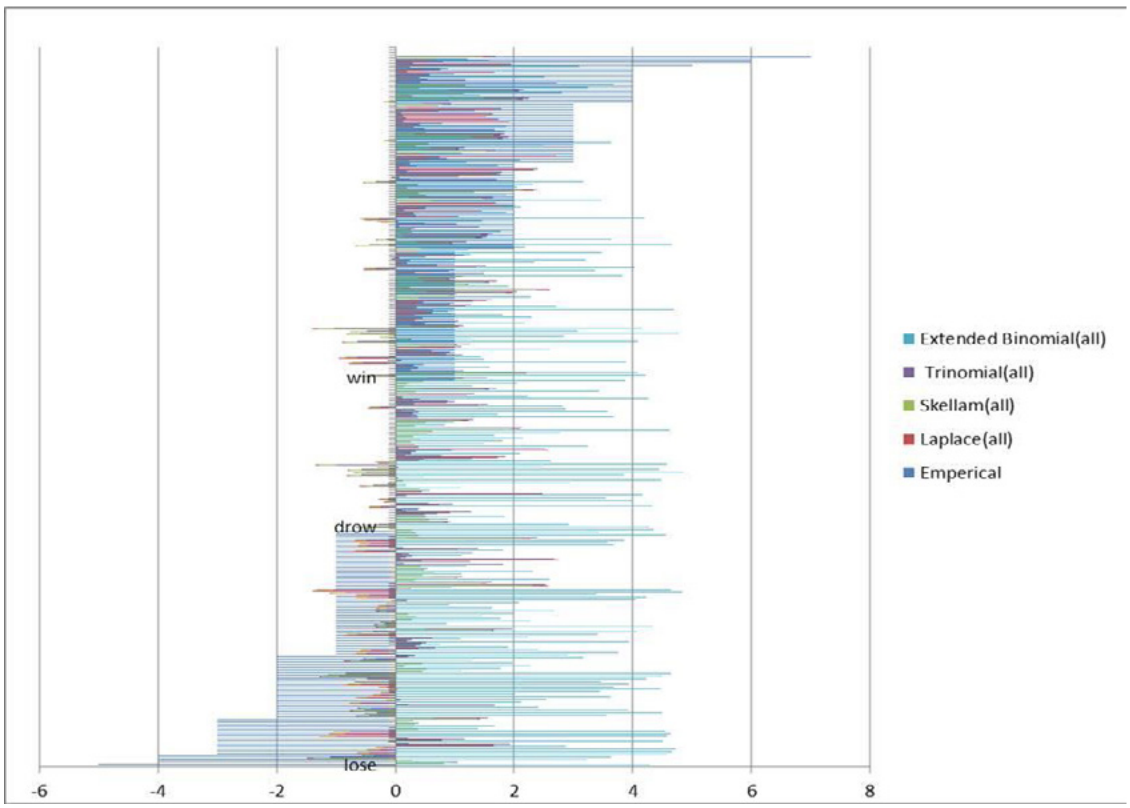

Fig. 1: Empirical versus the fitted values of the different models (all parameters).

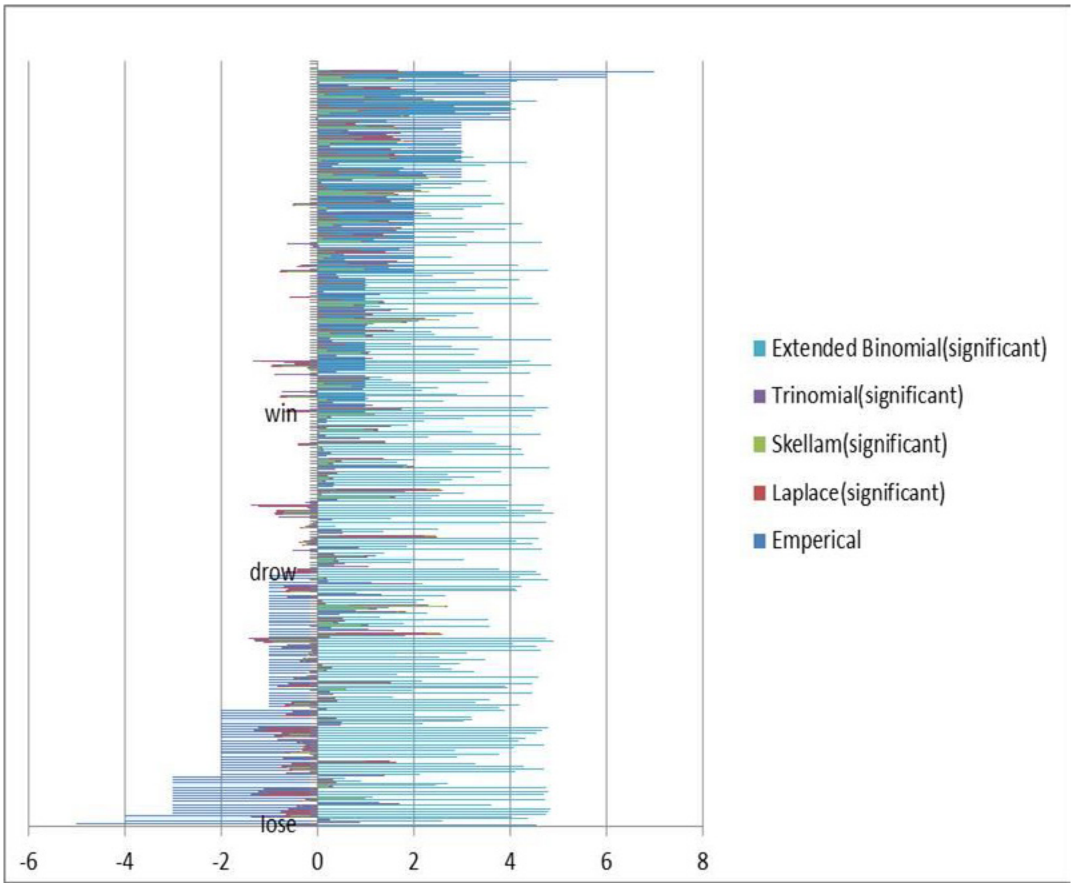

Fig. 2: Empirical versus the fitted values of the different models (significant).

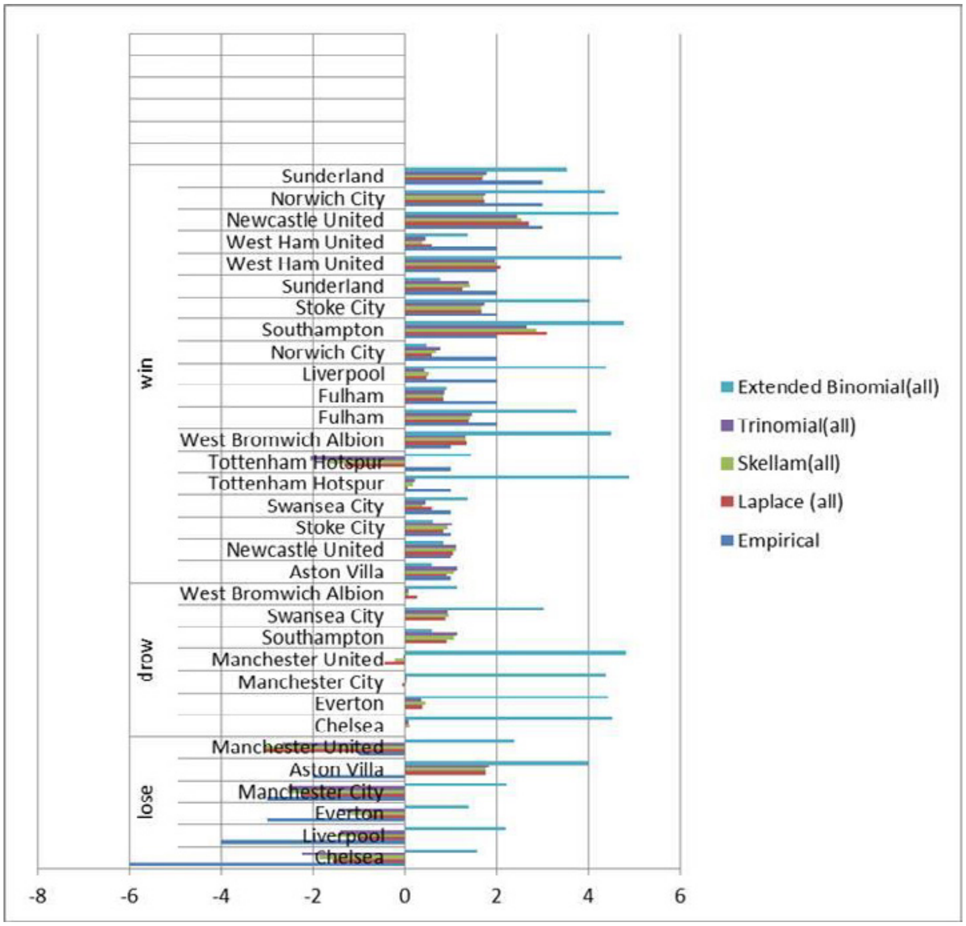

Fig 3: Empirical versus the fitted values of the different models (all parameters)

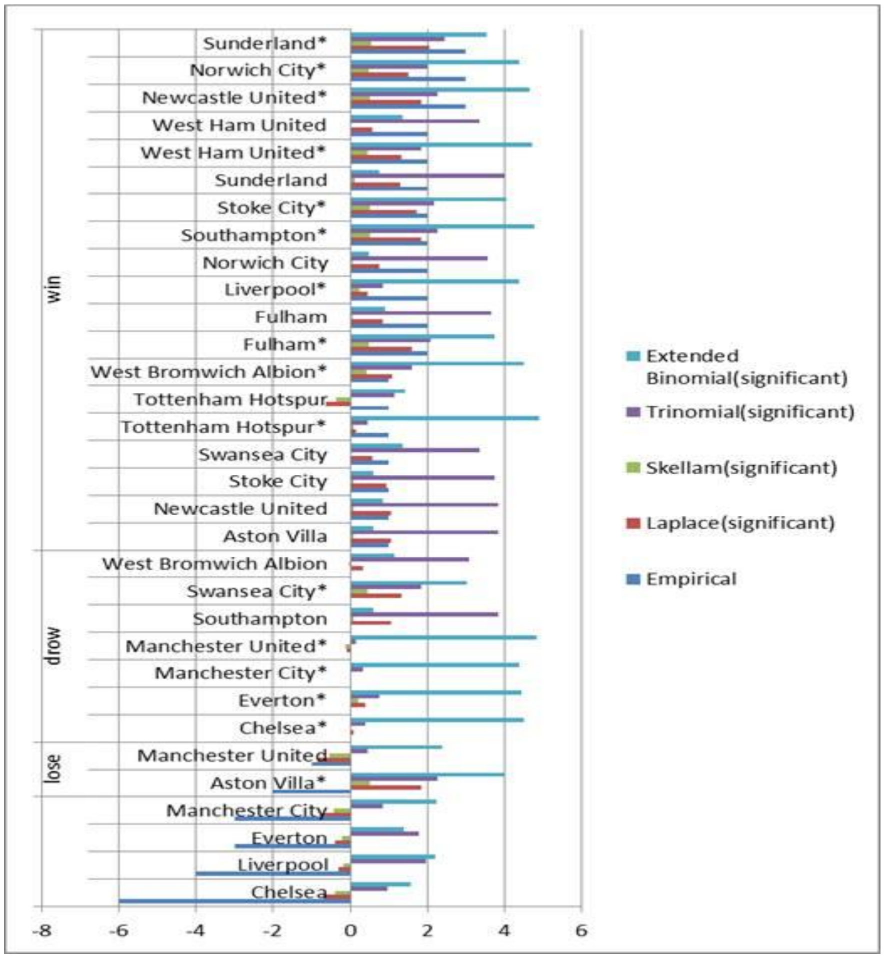

Fig 4: Empirical versus the fitted values of the different models (significant).

REFERENCES

ALZAID, A. & OMAIR, M. (2010) On the Poisson difference distribution inference and applications. Bulletin of the Malaysian Mathematical Sciences Society, 8(33), 17-45.

ALZAID, A. & OMAIR, M. (2012) An extended binomial distribution with applications. Communications in Statistics-Theory and Methods, 41(19), 3511-3527.

INUSAH, S. & KOZUBOWSKI, T. (2006) A discrete analogue of the Laplace distribution. Journal of statistical planning and inference, 136, 1090-1102.

KARLIS, D. & NTZOUFRAS, I. (2006) Bayesian analysis of the differences of count data. Statistics in medicine, 25, 1885-1905.

KARLIS, D. & NTZOUFRAS, I. (2008) Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. IMA Journal of Management Mathematics, 20(2), 133-145.

MCCULLAGH, P. & NELDER, J. (1989) Generalized Linear Models, Second ed., Monograph on Statistics and Applied Probability.

HILBE, J. (2011) Negative binomial regression, Second ed., Cambridge University Press, UK.

NELDER, J. & WEDERBURN, R. (1972) Generalized linear models. Journal of the Royal Statistical Society, A, 135, 370-384.

OMAIR, M., AlZAID, A. & ODHAH, O. (2016) A trinomial difference distribution. Revista Colombiana de Estadística, 39 (1), 1-15.

ONG, S., SHIMIZU, K. & MIN NG, C. (2008) A class of discrete distributions arising from difference of two random variables', Computational Statistics & Data Analysis 52(3), 1490-1499.

YEE, T. (2015) Vector generalized linear and additive models, New York, Springer.