Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v10n3p677

The effect of uncertainty on the assessment of individual performance: empirical evidence from professional soccer

By Iannario et al.

Published: 15 November 2017

Electronic Journal of Applied Statistical Analysis
Vol. 10, Issue 03, November 2017, 677-692
DOI: 10.1285/i20705948v10n3p677

The effect of uncertainty on the assessment of individual performance: empirical evidence from professional soccer

Fabio Fin[a], Maria Iannario*[b], Rosaria Simone[b], and Domenico Piccolo[b]

[b] *Department of Political Sciences, University of Naples Federico II*
Via L. Rodinò, 22, 80138, Naples, Italy

Published: 15 November 2017

The most important variables for measuring achievements in team sports such as soccer are physical condition, technical and tactical performance. Furthermore, in the last decades it has been noted that also real market value and reputation play a major role. Because of the complexity of the game, it is difficult to ascertain the relative importance of each of these variables and to establish an objective ranking of the best sportsmen. The aim of the present study is to develop a statistical model for analyzing the subjective evaluations expressed by raters on soccer players by stressing the effect of *uncertainty* and *don't know* responses in the decision-making process.

Keywords: Ordinal Response Model, Ratings, Uncertainty, CUB models, Don't Know Responses.

1 Introduction

Association football (soccer in the US) is one of the most spectator sports discussed for fun or for making money in betting markets. Recently, an increasing number of statistical models has been proposed to pursue game and market values predictions (Maher, 1982; Dobson and Goddard, 2011).

*Corresponding author: maria.iannario@unina.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

Individual's performance may be difficult to ascertain and judge for team sports like football. Indeed, different people have different opinions on player's technical capabilities: physical condition, technical and tactical performance, real market value measured by *the transfer fee* of the player (among others), as well as the responsibilities of each game role, may influence the *evaluation*.

In this paper we analyze the votes on the 2015 best men football player by comparing experts' judgments (journalists, coaches and team captains overseen and monitored by PricewaterhouseCoopers Switzerland -PwC- for FIFA *Ballon d'Or*) with the evaluations expressed by laymen over a rating scale with $m = 5$ categories. The performance of the players are assessed in both the classifications. Specifically, we select 11 top players among the 23 presented in the *Ballon d'Or* shortlist, chosen considering several factors: the fact that the forward players are more visible than those engaged in other roles, the team and the national leagues proportionally to the number of candidates (4 from Spain, 3 from Germany, 2 from England and 1 from Italy and France, respectively). In addition, the selection considers the soccer players who best qualified in different sport events in 2015 according to competition rules.

The idea is to measure the difference in the evaluations and in the consequent classifications by using a statistical model introduced for ratings (Piccolo, 2003) which manages both the *feeling* expressed on the soccer's performance and the *uncertainty* in the response process. We also consider the *don't know* (*dk*) choice in the selection of the ratings for each soccer (Manisera and Zuccolotto, 2014) and possible subjects' and objects' covariates which affect the evaluation (Piccolo and D'Elia, 2008; Iannario et al., 2017). In this regard, we consider both characteristics of respondents (gender, age, country, having a team membership or a *pay tv* subscription) and player's information (personal, market and performance details).

Data are taken from Transfermarkt (October, 26, 2015), WhoScored, European Football Database, the Guardian rank, among others, and from an observational study for which 350 respondents have been interviewed in 2016. The questionnaire has been administered through a web link (expiring after filling out the form) delivered to respondents on a voluntary basis in May 2016.

In this paper the focus is on the role that *uncertainty* plays in the classification and the extent it enriches the information concerning the final evaluation of the best players. The discussion will point out the differences in the measurements among clusters, the possible presence of overdispersion and the increasing amount of heterogeneity when the *dk* option is listed among the alternatives. It is worth to underline that the analysis here fulfilled can be applied to other sports.

The paper is organized as follows. Section 2 is devoted to the specification of the method and of the setting. Section 3 presents the case study, with the description of the data and the implementation of the models. Section 4 concludes with some summarizing comments on the proposed analysis and on derived results. An appendix outlines some inferential issues.

2 Method and setting

We consider a latent trait Y measured on a sample of n subjects asked to express an evaluation on a given ordinal scale from 1 to m, with higher values of the rating corresponding to higher levels of the latent trait. Let R be the random variable on the support $\{1, 2, \dots, m\}$ collecting the scores.

We assume that the population is divided into two groups according to a dichotomous variable A, independent of the latent trait, taking values 0 or 1 with probability ϕ and $1 - \phi$, respectively, and indicating whether a respondent is able ($A = 0$) or not ($A = 1$) to formulate the requested rating R. In this second case he/she selects the *don't know* alternative (*dk* for short). Thus, for a given subject $i = 1, \dots, n$, we set $a_i = 1 \Leftrightarrow I_i = 1$, where a_i is the value of A for subject i and I_i is the indicator function assuming value 1 if the subject i marks the *dk* option, so that $n_{dk} = \sum_{i=1}^{n} I_i$ denotes the total number of *dk* responses.

Conditioning on $A = 0$, the probability distribution of R, $Pr(R = r|A = 0; \boldsymbol{\theta}_0)$, $r = 1, \dots, m$, depends on a parameter vector $\boldsymbol{\theta}_0$ characterizing all the subjects with $a_i = 0$. This probability distribution follows some statistical model M_0 that can be fitted to the data after list-wise deletion of *dk* responses.

For $P(R = r|A = 1)$ a discrete Uniform distribution U_m on the support $\{1, 2, \dots, m\}$ is assumed. Generally, other alternative models are equally adequate for people who select *dk* option; however, the proposed choice may be derived by guesswork as motivated in Manisera and Zuccolotto (2016), pag. 105.

Thus, the marginal distribution of the response R is specified as the mixture model:

$$Pr(R; \boldsymbol{\theta}) = \phi Pr(R = r|A = 0; \boldsymbol{\theta}_0) + (1 - \phi)\,\frac{1}{m}\,, \tag{1}$$

where $\boldsymbol{\theta}' = (\phi, \boldsymbol{\theta}_0)'$. From the inferential point of view, given a sample of n subjects, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$, where $\mathbf{x}_i = (r_i, a_i)$, the log-likelihood function is given by:

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{x}) &= \sum_{i=1}^{n} \log[Pr(R = r_i, A = a_i; \boldsymbol{\theta})] \\ &= \sum_{i=1}^{n} (1 - I_i) \log[\phi\; Pr(R = r_i|A = 0; \boldsymbol{\theta}_0)] + \sum_{i=1}^{n} I_i \log\left(\frac{1-\phi}{m}\right) \\ &= \ell_1(\boldsymbol{\theta}_0) + \ell_2(\phi) \end{aligned} \tag{2}$$

where

$$\begin{aligned} \ell_1(\boldsymbol{\theta}_0) &= \sum_{i=1}^{n} (1 - I_i) \log[\,Pr(R = r_i|A = 0; \boldsymbol{\theta}_0)] \\ \ell_2(\phi) &= (n - n_{dk}) \log(\phi) + n_{dk} \log(1 - \phi) - n_{dk} \log(m). \end{aligned} \tag{3}$$

Because of this formulation, the parameter vector $\boldsymbol{\theta}_0$ is estimated by fitting the model M_0 assumed for the sample ratings after list-wise deletion of the *dk* responses, whereas the maximum likelihood estimator of ϕ turns out to be the relative frequency of the provided ratings:

$$\hat{\phi} = \frac{n - n_{dk}}{n}.$$

Following the motivations supplied in Manisera and Zuccolotto (2014), we assume for M_0 a CUB mixture (D'Elia and Piccolo, 2005) in which two latent components, named as *feeling* and *uncertainty*, express the level of agreement/satisfaction (or disagreement/disliking) with the item being evaluated and the indecision surrounding the discrete choice on the rating scale, respectively. In this case,

$$Pr(R = r | A = 0; \boldsymbol{\theta}_0) = \pi_0 \, b_r(\xi_0) + \frac{1 - \pi_0}{m}, \qquad r = 1, \dots, m; \tag{4}$$

where $b_r(\xi_0)$ denotes the shifted Binomial distribution with parameter ξ_0:

$$b_r(\xi_0) = \binom{m-1}{r-1} \xi_0^{m-r} (1 - \xi_0)^{r-1}, \; r = 1, 2, \dots, m.$$

Thus, the feeling (in the sense of preference) is measured by $1 - \xi_0$ and the uncertainty is weighted by $1 - \pi_0$. In this case, the marginal distribution in (1) of all responses also follows a CUB model with parameters $\boldsymbol{\theta}_0 = (\phi \, \pi_0, \, \xi_0)'$ (CUB$_{dk}$ for short). In particular, a higher uncertainty equal to $1 - \phi \, \pi_0$ is determined when considering the *dk* responses with respect to that specified for M_0 equal to $1 - \pi_0$. This variation is strictly related to the percentage of *dk* choices: thus, the *dk* effect is modelled on the expressed ratings by coherently estimating an increased indecision about the latent trait.

Alternatively, the CUB mixture (Iannario, 2014, 2015) extends the CUB framework allowing to deal with possible overdispersion: this is realized by considering a shifted Beta-Binomial distribution for the feeling component. The presence of *dk* responses affects the uncertainty parameter by systematically lowering it also when overdispersion is observed. Specifically, when we assume a CUB model with parameters $\boldsymbol{\theta}_0' = (\pi_0, \xi_0, \delta_0)'$ for $Pr(R = r | A = 0; \boldsymbol{\theta}_0)$ (where δ_0 denotes the overdispersion parameter), the marginal distribution in (1) follows a CUB model with parameters $\boldsymbol{\theta}_0 = (\phi \, \pi_0, \xi_0, \delta_0)'$. This possible extension is further discussed and tested in Section 3.3. More generally, different preference models can be chosen for $Pr(R = r | A = 0; \boldsymbol{\theta}_0)$ according to the approach proposed in Tutz et al. (2016) and Iannario and Piccolo (2016), for instance.

In this paper, we further investigate the effect of *dk* responses on uncertainty for clusters of respondents determined by dichotomous covariates. For a dummy covariate $\mathcal{D}$, the probabilities $Pr(A = 0 | \mathcal{D} = d)$ and $Pr(A = 1 | \mathcal{D} = d)$, $d = 0, 1$, may be estimated by the relative frequencies of expressed ratings and *dk* responses, respectively, conditional to $\mathcal{D} = d$. When there is no risk of confusion, we use the symbols $n^{(\mathcal{D}=d)}$ and $n_{dk}^{(\mathcal{D}=d)}$ to indicate the frequency of respondents having expressed a rating or chosen the *dk* option, respectively, such that $\mathcal{D} = d$. The methodology here presented and aimed to emphasize the role of the uncertainty parameter in presence of *dk* responses has been extended in case of nominal, ordinal or numerical covariates (Iannario et al., 2017).

Thus, we obtain a model for R - conditional to $\mathcal{D}$ - by considering two clusters of respondents determined by the link: $logit(\phi_i) = \omega_0 + \omega_1 d_i$. Then, given a sample of n subjects, $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_n)'$, where $\mathbf{s}_i = (r_i, a_i, d_i)$, the log-likelihood function is given by:

$$\begin{aligned}
\ell(\boldsymbol{\theta}|\mathbf{s}) &= \sum_{i=1}^{n} \log[Pr(R = r_i, A = a_i|\mathcal{D} = d_i; \boldsymbol{\theta})] \\
&= \sum_{i=1}^{n} \log[Pr(R = r_i|A = a_i, \mathcal{D} = d_i; \boldsymbol{\theta}) Pr(A = a_i|\mathcal{D} = d_i; \boldsymbol{\omega})] \\
&= \sum_{i=1}^{n}(1 - I_i) \log[\phi_i \, Pr(R = r_i|A = 0, \mathcal{D} = d_i; \boldsymbol{\theta}_0)] + \sum_{i=1}^{n} I_i \log\left(\frac{1-\phi_i}{m}\right) \\
&= \ell_1(\boldsymbol{\theta}_0) + \ell_2(\boldsymbol{\omega})
\end{aligned}$$

where we set $\boldsymbol{\omega} = (\omega_0, \omega_1)'$ and:

$$\begin{aligned}
\ell_1(\boldsymbol{\theta}_0) &= \sum_{i=1}^{n}(1 - I_i) \log[Pr(R = r_i|A = 0, \mathcal{D} = d_i; \theta_0)] \\
\ell_2(\boldsymbol{\omega}) &= \sum_{i=1}^{n}(1 - I_i) \, log(\phi_i) + \sum_{i=1}^{n} I_i \, log(1 - \phi_i) - n_{dk} log(m) \\
&= \sum_{i=1}^{n} log(\phi_i) - \omega_0 \, n_{dk} - \omega_1 n_{dk}^{(\mathcal{D}=1)} - n_{dk} log(m). \qquad (5)
\end{aligned}$$

Thus, $\boldsymbol{\theta}_0$ is estimated by maximizing the function $\ell_1(\boldsymbol{\theta}_0)$ depending on the selected model M_0 fitted to the rating data after list-wise deletion of dk responses.
Given the structure of the model, the mixing proportion ϕ_i is the probability that the i-th subject provides a rating: thus, by solving the likelihood equations for $\ell_2(\boldsymbol{\omega})$, the maximum likelihood estimate for the intercept coefficient is given by the log odds of expressing a rating for respondents with $d_i = 0$:

$$\hat{\omega}_0 = log\left(\frac{n^{(\mathcal{D}=0)} - n_{dk}^{(\mathcal{D}=0)}}{n_{dk}^{(\mathcal{D}=0)}}\right),$$

whereas the estimate of the regression coefficient for $\mathcal{D}$ is the log odds-ratio of the cross-classification of $\mathcal{D}$ and A, the indicator of having or not opted for the dk option (Iannario et al., 2017):

$$\hat{\omega}_1 = log\left(\frac{n_{dk}^{(\mathcal{D}=0)}}{n^{(\mathcal{D}=0)} - n_{dk}^{(\mathcal{D}=0)}}\right) - log\left(\frac{n_{dk}^{(\mathcal{D}=1)}}{n^{(\mathcal{D}=1)} - n_{dk}^{(\mathcal{D}=1)}}\right).$$

From an empirical point of view, interesting results may be obtained if $\mathcal{D}$ is significant also to explain the uncertainty parameter π: as customarily in the CUB models framework, this circumstance is investigated by using a logit link between π and $\mathcal{D}$: $logit(\pi_i) = \beta_0 + \beta_1 \, d_i$, so that the parameter vector $\boldsymbol{\theta}_0 = (\beta_0, \beta_1, \xi_0)'$ is estimated by fitting the CUB model assumed for the expressed ratings to sample data with a dummy covariate for the uncertainty parameter. Thus, in the simplified case of CUB mixtures, the marginal distribution of R is given by:

$$\begin{aligned} Pr(R=r|\mathcal{D}=d;\boldsymbol{\theta}) &= \phi_i\, Pr(R=r|A=0,\mathcal{D}=d;\boldsymbol{\theta}_0) + \frac{1-\phi_i}{m} \\ &= \phi_i \left[\pi_i b_r(\xi_0) + \frac{1-\pi_i}{m}\right] + \frac{1-\phi_i}{m} \\ &= \pi_i^{adj} b_r(\xi_0) + \frac{1-\pi_i^{adj}}{m}, \end{aligned} \tag{6}$$

where $\pi_i^{adj} = \phi_i\,\pi_i$. This means that the marginal distribution for the overall response R is a CUB model with feeling parameter ξ_0 and uncertainty parameter (depending on $\mathcal{D}$) given by π_i^{adj} (which is function of β_0, β_1 and ϕ_i). In other words, the *dk* effect on the uncertainty of the whole population results in an increase of the uncertainty itself also when fitting M_0 separately on each group of respondents. Analogous conclusions can be derived when M_0 is assumed to follow a CUB distribution: in this case, the parameter vector to be estimated is $\boldsymbol{\theta}_0 = (\beta_0, \beta_1, \xi_0, \delta_0)'$ (see Section 3.3 for an example). All the inferential issues about CUB and CUBE models are tackled by maximum likelihood estimation according to Piccolo (2006); Iannario (2014). In the Appendix we outline the computation of the information matrix for the proposed model. All the analysis has been run in R: the devoted code is available from Authors under request and it includes the usage of Package 'CUB' for CUB models fitting and testing (Iannario *et al.*, 2017).

3 Case study

The following section is devoted to the case study: after the description of the data, the modelling approach introduced in Section 2 is applied along with an extension of CUB models dealing with objects' (players) covariates. In conclusion, an example of generalization to CUB models with *dk* responses is presented.

3.1 Data description

For the subsequent twofold analysis, data are taken from different sources: specifically, we refer to:

- data concerning players' personal information (name, team, age, height, weight), market information (transfer fee, former team, duration of the contract, time in which the player joined the team) and performance information (on pitch time, actions at the ball, fouls, scores) collected from Transfermarkt, WhoScored, European Football Database and the Guardian rank;
- final ratings expressed in January 2016 from the *Ballon d'Or* team of experts;

- data arising from an observational study and collected in May 2016.

The last source of data involved 350 respondents who participated in the survey on a voluntary basis. The questionnaire was administered through a web link expired after filling out the form. Respondents were asked to express, on a scale from 1 to 5, their opinion on 11 top players selected from the 23 presented in the *Ballon d'Or* shortlist (Table 1). Specifically, they answered to the following item:

Indicate, on the basis of performance and overall behavior on and off the soccer pitch, if the following players deserve the Ballon d'Or 2015. The rating scale ranges from a minimum of 1 (no merit) to a maximum of 5 (full-on).

The *dk* option was available in the questionnaire and the percentage of *dk*'s per player is reported in Table 1, which also indicates the modal values and the (normalized) Laakso and Taagepera index $\mathcal{LT}$ (Laakso and Taagepera, 1979) as descriptive measures of location and heterogeneity of the distribution, respectively. In particular, the chosen heterogeneity index is defined by:

$$\mathcal{LT} = \frac{1}{m-1}\left[\left(\sum_{r=1}^{m} Pr(R=r;\boldsymbol{\theta})^2\right)^{-1} - 1\right].$$

This quantity is related to the Gini heterogeneity index via a one-to-one correspondence, but it is in general more convenient since its range is larger in common situations (Capecchi and Iannario, 2016). The last two columns in Table 1 refer to the final ranking expressed by experts (*Ballon d'Or*) and to the classification obtained from the observational study (*Evaluation*) according to the estimated feeling $1 - \hat{\xi}_0$, respectively.

Table 1: Summary results concerning the 11 players

Player	*dk*	$\mathcal{LT}$-index	*Mode*	*Ballon d'Or*	*Evaluation*
Messi	0.046	0.440	5	1	1
Cristiano R.	0.057	0.702	5	2	2
Neymar	0.031	0.725	4	3	3
Neuer	0.086	0.977	3	7	5
Hazard	0.117	0.844	1	8	9
Zlatan	0.094	0.974	3	11	4
Alexis	0.068	0.790	2	10	10
Vidal	0.065	0.703	2	17	11
Pogba	0.071	0.942	3	14	8
Rakitic	0.128	0.912	3	23	7
Lewandowski	0.114	0.970	3	4	6

The sample is mainly composed of males (71.43%), the average age is 31 years, with a standard deviation of 9.34 years, the education level concerns mostly (45.71%) high school diploma, followed by degree (37.43%). The 81.43% of respondents prefers to watch football matches on TV and about 67.71% of them has a *pay tv* subscription.

Interviewees who have or had a membership for a professional or amateur soccer team represent about 1/3 (65.43% no, 34.57% yes); whereas the 56.57% agrees with the evaluation system related to *Ballon d'Or*. CUB models fitting procedure can be intuitively represented by plotting estimated uncertainty against estimated feeling as a point in the parameter space. Based on this feature, Figure 1 shows the ranks of the expressed evaluations on the 11 players in the parameter space of CUB models before and after the adjustment for the *dk*'s. For each candidate player, the tail of the arrow corresponds to the estimated uncertainty (*abscissa*) and feeling (*ordinate*) obtained by fitting the model (without covariates) after list-wise deletion of *dk*'s. These start points reflect the position of the evaluations (Table 1) with the additional information on the level of uncertainty in the responses. Then, when the *dk*'s are taken into account, the uncertainty estimate of each item increases to reach the arrow head. The adjustment is very important for the comparison of different statements. For example, before adjustment, the uncertainty estimates for Hazard and Vidal are pretty close, but the different percentages of *dk* responses lead to a larger difference between adjusted uncertainty estimates. As expected, the feeling estimate is not affected by the adjustment.

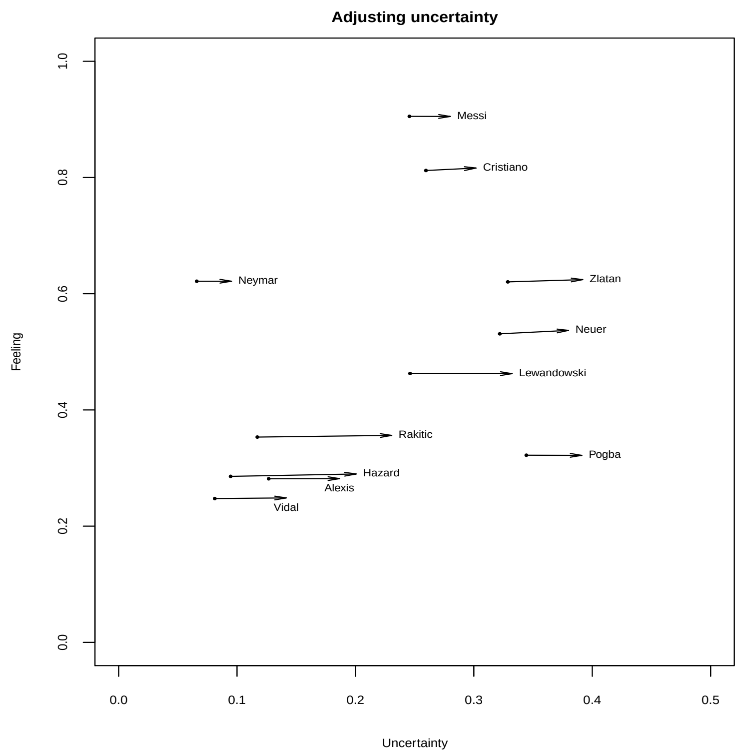

Figure 1: Uncertainty and feeling estimates for each player after list-wise deletion of *dk* responses (starting point of the arrows) and corresponding adjustment of the estimates for the *dk* responses (black arrows)

3.2 CUB models with dk option and covariates

In this subsection we analyze the responses to the sample survey on the performances of the selected 11 players by fitting data with CUB models, with both subjects' and objects' covariates.

Here, we consider the *dk* responses in two different ways: first, we take into account *all* items (players) simultaneously, whereas in the second case we analyze the *single* items separately. Specifically the analysis is carried out following two different strategies: **i**) including the percentage of *dk* responses among the objects' covariates in order to use it as inverse indicator for the top player celebrity (*knowledge*), and **ii**) applying the mixture model presented in Section 2 with the inclusion of dummy covariates. In more details:

i) We consider -as objects' covariates for CUB models- a score for each *player performance* provided by Who Scored which summarizes the number of goals, successful passes, dribbling skills, free kick goal rates, number of goal assists, successful saves, etc. (results concern 2014/2015 UEFA Champions League) and the *economic value* of the soccer player, taken from Transfermarkt, computed as the sum of the cost of the card and the gross salary of the player.

The aim of this analysis is to disclose characteristics of subjects who have expressed a rating, thus emulating the approach pursued in Iannario et al. (2017). In details, we seek to disentangle respondents and players characteristics possibly affecting the overall feeling and uncertainty about subjects' agreement with the ranking of the 11 players. Table 2 summarizes the result by reporting the estimated parameters and the corresponding standard errors (objects' covariates are starred).

Significant effects are observed both on uncertainty and feeling components; specifically, top player *knowledge**, *performance** and *gender* influence the level of uncertainty whereas preferences are influenced by top player *knowledge** and the *economic value** of each soccer player, the *agreement* with the evaluation system related to *Ballon d'Or*, the usual watching (*visualization*) of UEFA Champions League (UCL) matches, *gender* and the deviation from the mean of the logarithm of *age* in years. Specifically, if players are enumerated according to the ordering they are listed with in Table 1 (so that $j = 1$ corresponds to Messi, $j = 4$ corresponds to Neuer, and so on), the estimated CUB, regression model is given by:

$$\begin{aligned} logit(1-\pi_{ij}) &= 7.239 - 1.264\, gender_i + 66.481\, knowledge^*_j - 4.384\, performance^*_j \\ logit(1-\xi_{ij}) &= -11.082 + 0.304\, gender_i - 0.301\, age_i + 0.273\, agreement_i \\ &\quad +0.222\, visualization_i - 11.018\, knowledge^*_j + 2.289\, ec.\, value^*_j, \end{aligned} \tag{7}$$

for $j = 1, 2, \dots, 11$, $i = 1, 2, \dots, n$.

The interpretation derived from (7) implies that the uncertainty increases and the feeling decreases as the knowledge of the player lowers (high percentage of *dk*'s), and that *gender* affects both components: males (*gender*=1) are associated with lower uncertainty and higher feeling than females.

Table 2: CUB model on the evaluation of the best players, with subjects' and objects' covariates.

Components	Covariates	ML-estimates	Stand.errors	Wald-test
Uncertainty	Constant	−7.239	1.350	−5.363
	gender	1.264	0.403	3.132
	*knowledge**	−66.481	23.880	−2.784
	*performance**	4.384	0.780	5.618
Feeling	Constant	11.082	0.639	17.341
	gender	−0.304	0.073	−4.157
	age	0.301	0.104	2.900
	agreement	−0.273	0.060	−4.576
	visualization	−0.222	0.067	−3.311
	*knowledge**	11.018	1.275	8.639
	*economic value**	−2.289	0.116	−19.735
$\ell(\hat{\boldsymbol{\theta}})$	−5209.501	*BIC*	10508.9	

The level of uncertainty increases if the player' performance was not excellent. The feeling, instead, reduces among older people and increases with higher *visualization* of UCL and *agreement* with *Ballon d'Or* system of ratings.

ii) The second approach, partially introduced in the data source description, aims at highlighting different adjustments of the uncertainty estimates for distinct clusters. Here we report the difference in the evaluation of players induced by *gender*.

Table 3 summarizes the estimation results: π_F (π_M) and π_F^{adj} (π_M^{adj}) denote the estimated uncertainty parameter for females (males) without and with the adjustment, respectively), which are further represented in Figure 2. Generally, indecision is confirmed to be higher for females than males. Other analyses concerning, for instance, education, have been performed: respondents with degree are less uncertain in the responses, especially for the following players: Hazard, Zlatan, Alexis, Vidal, Lewandowski. The group of respondents having a *pay tv* subscription and that with a team membership are both characterized by a low uncertainty in assessing their evaluation.

3.3 CUBE models with dk option

By construction, the feeling component in CUB models is anchored to the shifted Binomial distribution, and thus it fails to account for possible extra-variability in preferences since its variance and mean value are mutually constrained. In such occurrence, there is the need to adequately specify the resulting *overdispersion* effect: within CUBE models framework, this task is accomplished by assigning a feeling measure that, on *a priori*

Table 3: Estimation of CUB and CUB_{dk} models with *gender* covariate on the uncertainty component for the 11 players

Player	π_F	π_M	ξ_0	π_F^{adj}	π_M^{adj}
Messi	0.568 *(0.090)*	0.822 *(0.046)*	0.095 *(0.014)*	0.505 *(0.097)*	0.805 *(0.050)*
Cristiano R.	0.492 *(0.114)*	0.814 *(0.053)*	0.184 *(0.018)*	0.437 *(0.116)*	0.785 *(0.060)*
Neymar	0.795 *(0.147)*	0.965 *(0.054)*	0.379 *(0.016)*	0.732 *(0.180)*	0.953 *(0.072)*
Neuer	0.340 *(0.193)*	0.772 *(0.088)*	0.463 *(0.023)*	0.296 *(0.182)*	0.720 *(0.102)*
Hazard	0.763 *(0.119)*	0.970 *(0.045)*	0.710 *(0.015)*	0.549 *(0.169)*	0.919 *(0.115)*
Zlatan	0.503 *(0.191)*	0.704 *(0.079)*	0.376 *(0.023)*	0.413 *(0.189)*	0.661 *(0.086)*
Alexis	0.854 *(0.104)*	0.880 *(0.057)*	0.718 *(0.016)*	0.674 *(0.187)*	0.870 *(0.062)*
Vidal	0.705 *(0.107)*	0.999 *(0.001)*	0.751 *(0.013)*	0.599 *(0.129)*	0.968 *(1.222)*
Pogba	0.662 *(0.138)*	0.652 *(0.092)*	0.678 *(0.024)*	0.556 *(0.156)*	0.627 *(0.096)*
Rakitic	0.760 *(0.140)*	0.925 *(0.064)*	0.644 *(0.017)*	0.585 *(0.191)*	0.843 *(0.123)*
Lewandowski	0.642 *(0.159)*	0.792 *(0.088)*	0.538 *(0.021)*	0.552 *(0.174)*	0.710 *(0.113)*

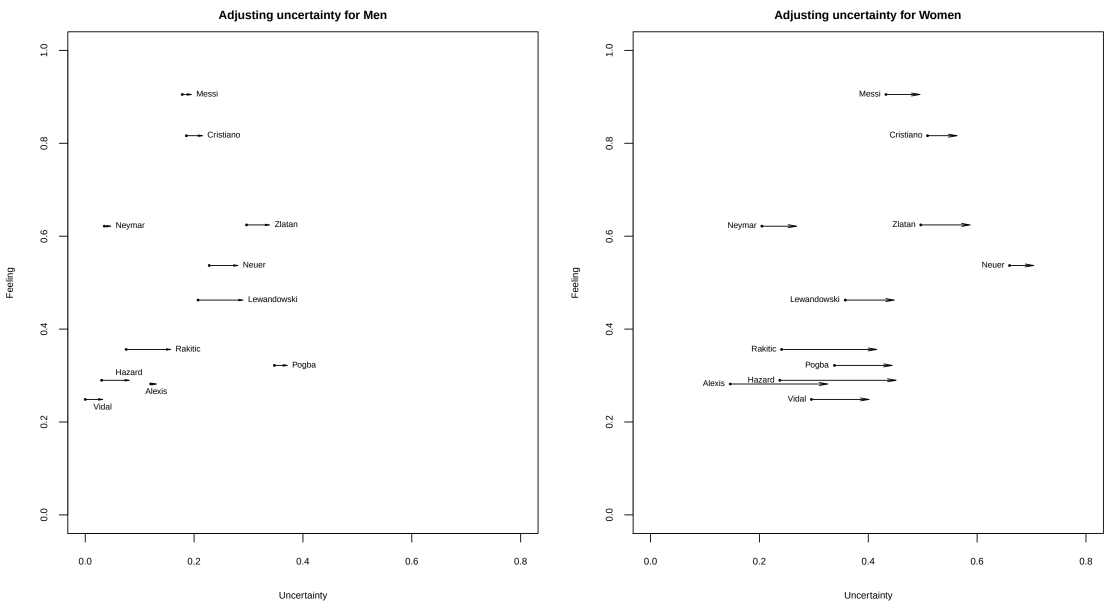

Figure 2: Uncertainty and feeling estimates for each player by *gender*

basis, is Beta distributed. This leads to the definition of CUBE models (Iannario, 2014), that are designed by shaping the feeling component with a shifted Beta-Binomial distribution (indeed, CUBE stands for *C*ombination of *U*niform and *BE*ta-Binomial):

$$be_r(\xi,\delta) = \binom{m-1}{r-1} \frac{\prod_{k=1}^{r}[1-\xi+\delta(k-1)] \prod_{k=1}^{m-r+1}[\xi+\delta(k-1)]}{[1-\xi+\delta(r-1)]\,[\xi+\delta(m-r)] \prod_{k=1}^{m-1}[1+\delta(k-1)]}, r = 1,\ldots,m. \quad (8)$$

This parametrization is mostly convenient especially to recognize that CUBE are nested into CUB models if $\delta = 0$. It indeed allows for effective models comparisons and selection in order to test the significance of the overdispersion effect (Iannario, 2015).

In this subsection we shortly discuss the implementation of CUBE models in presence of *dk* option focusing on Pogba, ranked in one of the last positions. Figure 3 summarizes the main results. We observe both a different performance in terms of *feeling*, with a

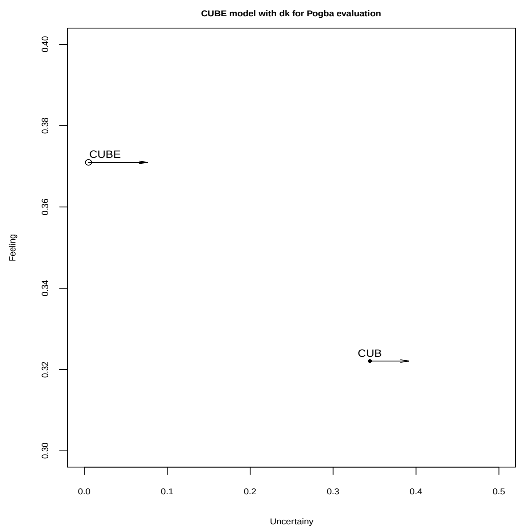

Figure 3: Effect of *dk* option on CUB and CUBE models

higher value estimated by the CUBE model, and a different assessment of uncertainty: in particular, we notice a significant confusion between indecision and overdispersion ($\hat{\delta} = 0.2$) when mis-specifying the overdispersion component. Moreover, the uncertainty measure in the responses is lower with CUBE models also when adjusting for the *dk* option. Summarizing, this modelling approach allows to distinguish between decrease in uncertainty due to overdispersion or to the *dk* effect.

4 Discussion and conclusion

In this paper we have analyzed the effect of uncertainty on the assessment of professional soccer individual performance with issues concerned with social and cultural aspects. From a statistical point of view, the data have been analyzed by means of a class of models able to detect two main features of respondents' perception (*feeling* and *uncertainty*), by further taking into account the information brought by the presence of *don't know* responses. The inclusion of the *dk* option adjusts the uncertainty parameter proportionally to the *dk* mixing proportion ϕ. Specifically, the presence of *dk* responses indicates difficulties towards the question: accordingly, the proposed modelling accounts for a higher indecision by adding more weight to the uncertainty in the data.

Results allow to understand the impact of items' complexity (the performance of soccer players) on the interviewees' evaluation and to take into account the increase in uncertainty due to *don't know* responses, both for the whole sample and for clusters of respondents. The approach can be suitably improved with an extended model to deal also with overdispersion.

A final consideration on the modelling of the *dk* responses in the class of CUB models derives from the following arguments. It is advisable to establish some rules that allow to assess if the specification of the *dk* option entails a significant variation in the uncertainty parameter. To this aim, one needs to merge the information on the *dk* effect directly within the sample of observed ratings (say, of size $n^\star = n - n_{dk}$). For instance, assume that a covariate $\mathcal{D}$ is found to be significant both for uncertainty and *don't know* responses, and that the latter relationship is explained via $logit(\hat{\phi}_i) = \hat{\omega}_0 + \hat{\omega}_1\, d_i$. Then, by adjusting for ϕ_i directly the regression coefficients in $logit(\hat{\pi}_i) = \hat{\beta}_0 + \hat{\beta}_1\, d_i$, one is lead to consider the model:

$$logit(\hat{\pi}_i^{\text{adj}}) = \hat{\beta}_0^{adj} + \hat{\beta}_1^{adj}\, d_i.$$

Thus, a model-selection criterion can be advocated to compare the fit performed by CUB$(\hat{\pi}_i, \hat{\xi})$ and CUB$(\hat{\pi}_i^{adj}, \hat{\xi})$ to the $n^\star$ data (as the BIC, for instance, or the mean log-likelihood) and to assess the impact of the *don't know* option as function of subjects' characteristics. An alternative check can be based on the Dissimilarity index computed on cluster-basis as in Manisera and Zuccolotto (2014).

In conclusion, results underline as the layman ranking of the best soccer players in 2015 relative to the official list reported in FIFA *Ballon d'Or* is confirmed only for the first positions and that the level of uncertainty in the responses increases for female and with no direct experience of soccer, low levels of education and lack of a *pay tv* subscription.

Acknowledgements: Research carried out in collaboration with the Big&Open Data Innovation Laboratory (BODaI-Lab), University of Brescia (project nr. 03-2016, title "Big Data Analytics in Sports", www.bodai.unibs.it/BDSports/), granted by Fondazione Cariplo and Regione Lombardia.

References

Capecchi, S. and Iannario, M. (2016). Gini heterogeneity index for detecting uncertainty in ordinal data. *Metron*, 74, 223–232.

D'Elia, A. and Piccolo, D. (2005). A mixture model for preference data analysis. *Computational Statistics & Data Analysis*, 49, 917–934.

Dobson, S. and Goddard, J. (2011). *The economics of football.* Cambridge University Press

Iannario, M. (2014). Modelling Uncertainty and Overdispersion in Ordinal Data. *Communications in Statistics. Theory and Methods*, 43, 771–786.

Iannario, M. (2015). Detecting latent components in ordinal data with overdispersion by means of a mixture distribution. *Quality & Quantity*, 49, 977–987.

Iannario, M., Manisera, M. and Zuccolotto, P. (2017). The treatment of don't know responses in the consumers' perceptions measurement: a survey in the agri-food sector. *Quality & Quantity*, 51, 765–778.

Iannario, M., Manisera, M., Piccolo, D. and Zuccolotto, P. (2017). Ordinal data models for no-opinion responses in attitude surveys. *Submitted.*

Iannario, M. and Piccolo, D. (2016). A comprehensive framework of regression models for ordinal data, *Metron*, **24**, 233–252.

Iannario, M., Piccolo, D. and Simone, R. (2017). CUB: A Class of Mixture Models for Ordinal Data (`R` package version 1.1.1). `http://CRAN.R-project.org/package=CUB`.

Laakso, M. and Taagepera, R. (1979). Effective number of parties: a measure with application to West Europe. *Comparative Political Studies*, 12, 3–27.

Maher, M.J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3), 109–118.

Manisera, M. and Zuccolotto, P. (2014). Modeling "Don't know" responses in rating scales. *Pattern Recognition Letters*, 45, 226–234.

Manisera, M. and Zuccolotto, P. (2016). Treatment of "don't know" responses in a mixture model for rating data. *Metron*, 74, 99–115.

Piccolo, D. (2003). On the moments of a mixture of uniform and shifted binomial random variables. *Quaderni di Statistica*, 5, 85–104.

Piccolo, D. (2006). Observed information matrix for MUB models. *Quaderni di Statistica*, 8, 33–78.

Piccolo, D. and D'Elia, A. (2008). A new approach for modelling consumers' preferences. *Food Quality and Preference*, 19(3), 247–259.

Tutz, G., M. Schneider, M. Iannario and D. Piccolo (2016). Mixture models for ordinal responses to account for uncertainty of choice. *Advances in Data Analysis and Classification*, 11(2), 281–305.

Appendix

Given the log-likelihood function (2) of the standard model without covariates, the parameters ϕ and $\boldsymbol{\theta}_0$ can be estimated separately by maximizing $\ell_2(\phi)$ and $\ell_1(\boldsymbol{\theta}_0)$, respectively. Let $\boldsymbol{\theta} = (\phi, \boldsymbol{\theta}_0)'$ and $\mathcal{I}_{\theta_i\,\theta_j} = -\left[\dfrac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log \ell(\boldsymbol{\theta}|x)\right], \forall i, j$, the observed information matrix is:

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = \left(\begin{array}{c|c} \mathcal{I}_{\phi\,\phi} & 0 \\ \hline 0 & \mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{CUB}}) \end{array}\right), \qquad \mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{CUB}}) = \begin{pmatrix} \mathcal{I}_{\pi\,\pi} & \mathcal{I}_{\pi\,\xi} \\ \mathcal{I}_{\xi\,\pi} & \mathcal{I}_{\xi\,\xi} \end{pmatrix}$$

where the sub-matrix $\mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{CUB}})$ is the information matrix for the CUB model (M_0). Since the novelty of the analysis discussed in the paper is the modelling of *dk* responses, the focus here is on $\mathcal{I}_{\phi\,\phi}$: the first and second derivatives of the log-likelihood (3) are:

$$\ell_2'(\phi) = \frac{(n - n_{dk})}{\phi} - \frac{n_{dk}}{1-\phi}, \qquad \ell_2''(\phi) = -\frac{(n - n_{dk})}{\phi^2} - \frac{n_{dk}}{(1-\phi)^2}.$$

For models with covariates, we restrict our attention on a dichotomous covariate (as in the paper). Also in this case the information matrix is of block type:

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = \left(\begin{array}{c|c} \mathcal{I}(\hat{\boldsymbol{\omega}}) & 0 \\ \hline 0 & \mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{CUB}}) \end{array}\right), \qquad \mathcal{I}(\hat{\boldsymbol{\omega}}) = \begin{pmatrix} \mathcal{I}_{\omega_0\,\omega_0} & \mathcal{I}_{\omega_0\,\omega_1} \\ \mathcal{I}_{\omega_1\,\omega_0} & \mathcal{I}_{\omega_1\,\omega_1} \end{pmatrix}$$

with the bottom-right block $\mathcal{I}(\hat{\boldsymbol{\theta}}_{\text{CUB}})$ corresponding to the information matrix of a CUB model with covariate on the uncertainty parameter (Piccolo, 2006). Then, having set $logit(\phi_i) = \omega_0 + \omega_1\,d_i$, one has:

$$\frac{\partial\phi_i}{\partial\omega_0} = \phi_i\,(1-\phi_i), \qquad \frac{\partial\phi_i}{\partial\omega_1} = d_i\,\phi_i\,(1-\phi_i),$$

with $\ell_2(\boldsymbol{\omega})$ given by (5). Thus, the entries of $\mathcal{I}(\hat{\boldsymbol{\omega}})$ are:

$$\begin{aligned} \mathcal{I}_{\omega_0\,\omega_0} &= \sum_{i=1}^{n} \phi_i\,(1-\phi_i) \\ \mathcal{I}_{\omega_0\,\omega_1} = \mathcal{I}_{\omega_1\,\omega_0} &= \sum_{i=1}^{n} d_i\,\phi_i\,(1-\phi_i) \\ \mathcal{I}_{\omega_1\,\omega_1} &= \sum_{i=1}^{n} d_i^2\,\phi_i\,(1-\phi_i) = \mathcal{I}_{\omega_0\,\omega_1}. \end{aligned}$$

Finally, the asymptotic variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\theta})$ of the ML estimator of $\boldsymbol{\theta}$, computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} = (\hat{\phi}, \hat{\boldsymbol{\theta}}_0)'$ is given by:

$$\boldsymbol{V}(\hat{\boldsymbol{\theta}}) = \left[\mathcal{I}(\hat{\boldsymbol{\theta}})\right]^{-1} .$$

from which (asymptotic) standard errors can be derived.