



MEASURING UNCERTAINTY OF OUTCOME: AN ANALYTIC APPROACH

KJETIL K. HAUGEN

Abstract. This paper proposes an alternative mathematical formulation to one (among many) ways of estimating uncertainty of outcome. This alternative formulation may make such estimation processes easier for both practitioners and researchers. Additionally, some approximative estimators¹ are derived, tested and discussed. Even though these estimators may not perform perfectly for football leagues, they may have a stronger potential for use in other sports.

1. INTRODUCTION

One important feature that makes sports economic theory different from traditional economic theory is the so called uncertainty of outcome (UO). This concept, originally introduced by Rottenberg [9], states that sports competitions need excitement to produce willingness to pay. If the crowd on a football match knew the outcome of the match beforehand, they would probably be more willing to spend their valuable time otherwise than freezing or getting wet at a stadium. As a consequence, much research has been done to test the so-called UO-hypothesis. That is, does UO actually affect demand in the real world? Some examples of this line of research may for instance be found in [1, 3, 8] Apart from a few exceptions, limited research has been done on discussing computational properties of various methods for estimating UO.

This paper discusses one of these estimating methods for the uncertainty of outcome (UO) measure, ρ_L . This measure was introduced in [5]. As opposed to other UO measures, this one opens up for some interesting analytic opportunities as it has a mathematical structure different from other typical measures [2]. I will show that this structure can be utilized in finding relatively easy computational rules as well as quite simple approximative versions, easing the computational burden for practitioners. At the same time, analytic properties such as comparative static analysis open up.

Section 2 presents the necessary definitions, and Section 3 performs the analytical reformulation of ρ_L . Section 4 and appendix B suggest various approximated estimators, while Section 5 is a conclusion.

MSC (2010): primary 62C99, 62F03, 91A80.

Keywords: uncertainty of outcome, numerical simplification, approximative estimation.

¹The reader should be aware of the fact that the term estimator has a slightly different meaning here than the 'normal' use in statistics.

2. NECESSARY DEFINITIONS AND EQUATIONS

The following definitions and equations can for instance be found in [4]²:

- LCP_i : Least competitive points score in some league table for team i ,
- MCP : Maximally competitive point score in some league (all matches, end in a draw),
- AP_i : Actual point score in some league for team i ,
- T : Number of teams in the league,
- N : Number of matches in the league,
- ω_p : Number of points received by a team for a victory,
- δ_p : Number of points received by a team for a draw,

$$\rho_L = \frac{\sum_{i=1}^T (LCP_i - AP_i)^2}{\sum_{i=1}^T (LCP_i - MCP)^2} \cdot 100, \quad (2.1)$$

$$LCP_i = (N - 2(i - 1))\omega_p, \forall i \in \{1, \dots, T\}. \quad (2.2)$$

For a normal double Round Robin tournament:

$$\begin{aligned} T &= \frac{N}{2} + 1, \\ MCP &= N\delta_p. \end{aligned} \quad (2.3)$$

3. COMPUTATIONAL REFORMULATION OF ρ_L

Let me start by looking at $LCP_i - MCP$ in the denominator in (2.1). Substituting the content from equations (2.2) and (2.3) I get:

$$LCP_i - MCP = (N - 2(i - 1))\omega_p - N\delta_p = [(N + 2)\omega_p - N\delta_p] - [2\omega_p] \cdot i$$

or

$$C_0 - C_1 \cdot i,$$

where $C_0 = [(N + 2)\omega_p - N\delta_p]$ and $C_1 = [2\omega_p]$. Obviously, “constants” C_0, C_1 depend on N, ω_p and δ_p but not on i , which is the relevant variable in the upcoming summation.

Then, $(C_0 - C_1 \cdot i)^2 = C_0^2 - 2C_0C_1 \cdot i + C_1^2 \cdot i^2$, and

$$\sum_{i=1}^T (LCP_i - MCP)^2 = C_0^2 T - C_0 C_1 T(T + 1) + \frac{C_1^2}{6} T(T + 1)(2T + 1) \quad (3.1)$$

by well known summation formulas $\sum_i^n i = \frac{n(n+1)}{2}$ and $\sum_i^n i^2 = \frac{n(n+1)(n+1)}{6}$.

Equation (3.1) can be slightly simplified; for example as:

$$\sum_{i=1}^T (LCP_i - MCP)^2 = T \left(C_0^2 - C_1(T + 1) \left[C_0 - \frac{C_1}{6}(2T + 1) \right] \right). \quad (3.2)$$

² A lost match gives 0 points, without loss of generality.

Now, $C_0 = [(N + 2)\omega_p - N\delta_p]$ and $C_1 = [2\omega_p]$ together with $T = \frac{N}{2} + 1$ can be substituted back into equation (3.2) to produce the (ugly-looking) expression:

$$\left(\frac{N}{2} + 1\right) \left([(N + 2)\omega_p - N\delta_p]^2 - 2\omega_p \left(\frac{N}{2} + 2\right) \left\{ (N + 2)\omega_p - N\delta_p - \frac{\omega_p}{3}(N + 3) \right\} \right). \quad (3.3)$$

Let me furthermore regard this expression as a function, and call it $f(N, \omega_p, \delta_p)$.

Now, in practice, (almost) all football leagues apply the 3-1-0 point score system. That is, $\omega_p = 3$ and $\delta_p = 1$. The number of teams in various leagues does however vary somewhat. Hence, it may be of interest to look at $f(N, 3, 1)$. I find (after some tedious algebra):

$$f(N, 3, 1) = N \left(\frac{N}{2} + 1\right) (N + 3). \quad (3.4)$$

Now, before moving on to the more insightful parts of this paper (see Section 4) let me investigate whether I can reformulate the original nominator expression for easier computations. The nominator is:

$$\sum_{i=1}^T (LCP_i - AP_i)^2 = \sum_{i=1}^T LCP_i^2 - 2 \sum_{i=1}^T LCP_i \cdot AP_i + \sum_{i=1}^T AP_i^2. \quad (3.5)$$

Taking the first part of the expression (3.5) and substituting from equation (2.2) yields:

$$\begin{aligned} \sum_{i=1}^T LCP_i^2 &= \sum_{i=1}^T [(N - 2(i - 1))\omega_p]^2 \\ &= \omega_p^2 \left[TN^2 - 4N \sum_{i=1}^T (i - 1) + 4 \sum_{i=1}^T (i - 1)^2 \right]. \end{aligned} \quad (3.6)$$

Now, summing the arithmetic series is straightforward:

$$\sum_{i=1}^T (i - 1) = \frac{T(T - 1)}{2} \quad \text{and} \quad \sum_{i=1}^T (i - 1)^2 = \frac{T(T - 1)(2T - 1)}{6}.$$

Then, (3.6) can, after some algebra, be expressed as:

$$\sum_{i=1}^T LCP_i^2 = \frac{1}{6}\omega_p^2 N(N + 1)(N + 2). \quad (3.7)$$

By defining the table point score average $\overline{AP} = \left(\frac{1}{T} \sum_{i=1}^T AP_i\right)$, the positional weighted point score average $\overline{\overline{AP}} = \left(\frac{1}{T} \sum_{i=1}^T i \cdot AP_i\right)$ and the squared point score average $\overline{AP^2} = \left(\frac{1}{T} \sum_{i=1}^T AP_i^2\right)$, the second sum in (3.5) is:

$$-2 \sum_{i=1}^T LCP_i \cdot AP_i = 2\omega_p (N + 2) \overline{\overline{AP}} - \omega_p (N + 2)^2 \overline{AP} \quad (3.8)$$

and the third sum is:

$$\sum_{i=1}^T AP_i^2 = \left(\frac{N}{2} + 1\right) \overline{AP}^2 = \frac{1}{2}(N+2)\overline{AP}^2. \quad (3.9)$$

Observing that $(N+2)$ is a common factor in all equations (3.7), (3.8) and (3.9), the final expression for the nominator is:

$$(N+2) \left[\frac{1}{6}\omega_p^2 N(N+1) + 2\omega_p \overline{\overline{AP}} - \omega_p(N+2)\overline{AP} + \frac{1}{2}\overline{AP}^2 \right]. \quad (3.10)$$

I call this function $g(N, \omega_p, \delta_p, \overline{AP}, \overline{\overline{AP}}, \overline{AP}^2)$. Then, our reformulation of ρ_L is available as (simply dividing equation (3.10) on (3.3)):

$$\rho_L = \frac{\frac{1}{6}\omega_p^2 N(N+1) + 2\omega_p \overline{\overline{AP}} - \omega_p(N+2)\overline{AP} + \frac{1}{2}\overline{AP}^2}{\frac{1}{2}[(N+2)\omega_p - N\delta_p]^2 - 2\omega_p \left(\frac{N}{2} + 2\right) \left\{ (N+2)\omega_p - N\delta_p - \frac{\omega_p}{3}(N+3) \right\}}. \quad (3.11)$$

Doing the same as in equation (3.4), finding:

$$g(N, 3, 1, \overline{AP}, \overline{\overline{AP}}, \overline{AP}^2) = (N+2) \left\{ \frac{3}{2}N(N+1) + 6\overline{\overline{AP}} - 3(N+2)\overline{AP} + \frac{1}{2}\overline{AP}^2 \right\}$$

and (for the given normal point score system $\omega_p = 3$ and $\delta_p = 1$):

$$\rho_L = \frac{3N(N+1) + 12\overline{\overline{AP}} - 6(N+2)\overline{AP} + \overline{AP}^2}{N(N+3)}. \quad (3.12)$$

An even “nicer looking” expression can be found if the number of matches played in the league is given. Most big European leagues such as Premier League (*PL*), Serie A (*SA*) and La liga (*LL*) contain 20 teams (T) or 38 matches (N). Given this setting, (3.12) simplifies to:

$$\rho_L^{PL,SA,LL} = \frac{4446 + 12\overline{\overline{AP}} - 240\overline{AP} + \overline{AP}^2}{1558}. \quad (3.13)$$

The Norwegian case (Eliteserien) looks slightly different – $T = 16$, $N = 30$

$$\rho_L^{Norway} = \frac{2790 + 12\overline{\overline{AP}} - 192\overline{AP} + \overline{AP}^2}{990}. \quad (3.14)$$

Some practitioners might find an expression in the form of equations (3.13), (3.14) easier to handle than the original expression (2.1). Still, it contains the terms \overline{AP} , $\overline{\overline{AP}}$, \overline{AP}^2 , which may be difficult for a football-practitioner without immediate access to a spreadsheet or calculator. As a consequence, the next part derives some practically feasible approximations or heuristics meant to provide a quick (though approximative) estimate on uncertainty of outcome.

4. PRACTITIONER APPROXIMATIONS FOR ρ_L

One way of progressing towards a quick (and dirty) approximation might be to realise that the average point score in a league (\overline{AP}) is a measure which practitioners relate to. As a consequence, if it is possible to develop an approximation involving only \overline{AP} excluding \overline{AP} and $\overline{AP^2}$, it may work well.

The following two candidates are investigated:

$$\overline{AP^2} = \frac{1}{T} \sum_{i=1}^T AP_i^2 \approx \left(\frac{1}{T} \sum_{i=1}^T AP_i \right)^2 = (\overline{AP})^2, \quad (4.1)$$

$$\overline{\overline{AP}} \approx \frac{T+1}{2} \overline{AP}. \quad (4.2)$$

It can be shown that the more balanced the league, the better the approximations. In a perfectly balanced or well balanced league, point score differences between teams on the final table are not very large. That is $AP_i \approx AP_j \forall \{i, j\}$. Given $AP_i = AP \forall i$ as the approximation, the true estimator;

$$\overline{AP^2} = \frac{1}{T} \sum_{i=1}^T AP_i^2 = \frac{1}{T} \sum_{i=1}^T AP^2 = \frac{1}{T} \cdot T \cdot AP^2 = AP^2$$

and the approximation from (4.1) is

$$\left(\frac{1}{T} \sum_{i=1}^T AP_i \right)^2 = \left(\frac{1}{T} \sum_{i=1}^T AP \right)^2 = \left(\frac{1}{T} \cdot T \cdot AP \right)^2 = AP^2.$$

A similar argument can be performed for the other approximation (4.2). The true estimator is:

$$\overline{\overline{AP}} = \frac{1}{T} \sum_{i=1}^T i \cdot AP_i = \frac{1}{T} \cdot AP \sum_{i=1}^T i = AP \cdot \frac{1}{T} \cdot \frac{T(T+1)}{2} = \frac{T+1}{2} \cdot AP$$

while the approximation from (4.2) is

$$\frac{T+1}{2} \overline{AP} = \frac{T+1}{2} \left(\frac{1}{T} \sum_{i=1}^T AP_i \right) = \frac{T+1}{2} \left(\frac{12}{T} \cdot T \cdot AP \right) = \frac{T+1}{2} \cdot AP.$$

As I have launched two different approximations in equations (4.1) and (4.2) it is of course also possible to enumerate over these options constructing maximally three approximative estimators for ρ_L . Let me define

- $\widehat{\rho}_L^{23}$: An approximative estimator using (4.1) only.
- $\widehat{\rho}_L^{24}$: An approximative estimator using (4.2) only.
- $\widehat{\rho}_L^{23 \wedge 24}$: An approximative estimator using both (4.1) and (4.2).

Given the above definitions, equation (3.12) can be used to obtain the corresponding approximative estimators:

$$\hat{\rho}_L^{23} = \frac{3N(N+1) + 12\overline{AP} + \overline{AP}(\overline{AP} - 6(N+2))}{N(N+3)}, \quad (4.3)$$

$$\hat{\rho}_L^{24} = \frac{3N(N+1 - \overline{AP}) + \overline{AP}^2}{N(N+3)}, \quad (4.4)$$

$$\hat{\rho}_L^{23 \wedge 24} = \frac{3N(N+1) + \overline{AP}(\overline{AP} - 3N)}{N(N+3)}. \quad (4.5)$$

An example of these approximations³ and their behaviour are given in Table 1⁴.

Table 1. Approximations for ρ_L in Norwegian *Eliteserien* – 2017.

	ρ_L	$\hat{\rho}_L^{23}$	$\hat{\rho}_L^{24}$	$\hat{\rho}_L^{23 \wedge 24}$
<i>Eliteserien</i>	38.78%	30.58%	86.85%	78.65%

As Table 1 indicates, our approximations do not perform very well. A $\hat{\rho}_L^{23}$ of 30.58% is not too far from the true value of 38.78%, but the other two suggestions differ too much. This is of course not very surprising. After all, our basic assumption of relatively equal point scores between teams on the final table are clearly far from being met in the data⁵. Actually, European football leagues are not as competitive as they used to be, see for instance [7].

However, US sports may be a more interesting candidate for applying these approximations. The closed leagues of US sports have a long tradition for regulations meant to improve competitive balance (or increasing ρ_L). Examples of such regulations are “reserve clause”, “reverse order drafts”, “salary caps” or “revenue sharing”⁶. As a consequence, logically, one should expect higher competitive balance in US sports leagues, and hence (hopefully) better quality of the given approximations. That is, even though the presented approximations may be less suitable for European football leagues, they might prove interesting for analysing (more competitive) US sports leagues.

If football leagues are less competitive, an alternative approach could be to construct estimators based on an inverse principle. That is, instead of an assumption such as $AP_i \approx AP_j$, one could assume that most leagues are highly incompetent (imbalanced) and assume alternatively that $AP_I \approx LCP_i$. Expressions similar to those above can be derived based on this principle. For completeness, they are added in appendix B.

Unfortunately, even if European football leagues are less competitive than they used to be, they are still too competitive for such an approximative scheme to function satisfactorily. However, these approximations may still have some value

³The Example is taken from from Norwegian *Eliteserien* for the 2016/17 season.

⁴Refer to appendix A for relevant data underlying the calculations.

⁵Refer to appendix A for the actual final table in *Eliteserien* 2017.

⁶Refer to [10] for a comprehensive discussion of these concepts.

in other sports. An interesting candidate is European handball, which shows remarkable examples of lack of uncertainty of outcome – refer to [6].

5. CONCLUSIONS

The main finding in this paper is perhaps equation (3.11). This formulation serves as an alternative to the definition (2.1). Hopefully, both practitioners and researchers may find this version easier both to work with analytically and numerically. The fact that most football leagues use the 3-1-0 point score system made it possible also to construct equations (3.12) as well as specific versions (3.13) and (3.14). As I see it, I do believe these expressions ought to make uncertainty of outcome estimates more available to the practitioner.

In addition, the paper proposes some approximations. As discussed in the paper, these approximations may not be perfectly suitable for most football leagues, but – as argued – they ought to be of interest for other sports such as handball and US team sports. As such, these approximations may be of interest to a growing family of both researchers and practitioners interested in a deeper empirical analysis of sports.

Finally, the illustrating computational results obtained in this paper show that the considered indicators and their approximations can be further studied as useful measures for the description of disbalance/balance of sport competitions. It can be especially valuable from the viewpoint of possible organizational changes motivated by the idea to keep or increase the interest of the audience. Such characteristics can also be further generalized to describe a group-wise disbalance of certain competitions.

APPENDIX A. FINAL TABLE *Eliteserien* – 2017

Table 2 contains the final table from *Eliteserien* in Norway from the 2017 season⁷. Some relevant data (calculated from Table 2) used in previous sections:

$$\begin{aligned}\overline{AP} &= 41.3125, \\ \overline{\overline{AP}} &= 311.5, \\ \overline{AP^2} &= 1787.9375.\end{aligned}$$

APPENDIX B. APPROXIMATIONS FOR IMBALANCED SPORTS LEAGUES

Given an initial assumption of $AP_i \approx LCP_i$, the following holds:

$$\overline{AP^2} = \frac{1}{T} \sum_{i=1}^T AP_i^2 \approx \sum_{i=1}^T LCP_i^2 = \frac{1}{T} \frac{1}{6} \omega_p^2 N(N+1)(N+2) = 3N(N+1) \quad (\text{B.1})$$

⁷M means matches played. W, D and L are matches won, drawn and lost respectively. +, – and +/- are goals scored, goals conceived and goal difference respectively, while P is number of points.

Table 2. Final table Norwegian *Eliteserien* – 2017.

	M	W	D	L	+	-	+/-	P
Rosenborg	30	18	7	5	57	20	+37	61
Molde	30	16	6	8	50	35	+15	54
Sarpsborg 08	30	13	12	5	50	36	+14	51
Strømsgodset	30	14	8	8	45	37	+8	50
Brann	30	13	8	9	51	36	+15	47
Odd	30	12	6	12	27	39	-12	42
Kristiansund BK	30	10	10	10	44	46	-2	40
Vålerenga	30	11	6	13	48	46	+2	39
Stabæk	30	10	9	11	46	50	-4	39
Haugesund	30	11	6	13	35	39	-4	39
Tromsø	30	10	8	12	42	49	-7	38
Lillestrøm	30	10	7	13	40	43	-3	37
Sandefjord	30	11	3	16	38	51	-13	36
Sogndal	30	8	8	14	38	48	-10	32
Aalesund	30	8	8	14	38	50	-12	32
Viking	30	6	6	18	33	57	-24	24

by equation (3.7). Furthermore (after a fair amount of algebra):

$$\overline{AP} = \frac{1}{T} \sum_{i=1}^T i \cdot AP_i \approx \frac{1}{T} \sum_{i=1}^T i \cdot LCP_i = \frac{1}{4}(N+4)(2N+3). \quad (\text{B.2})$$

Given approximations (B.1) and (B.2), expressions similar to those in equations (4.3)–(4.5) can be derived:

$$\begin{aligned} \widehat{\rho}_L^{23} &= \frac{6N(N+1) + 12\overline{AP} - 6(N+2)\overline{AP}}{N(N+3)}, \\ \widehat{\rho}_L^{24} &= \frac{3N(N+1) + 3(N+4)(2N+3) - 6(N+2)\overline{AP} + \overline{AP}^2}{N(N+3)}, \\ \widehat{\rho}_L^{23 \wedge 24} &= \frac{6N(N+1) + 3(N+4)(2N+3) - 6(N+2)\overline{AP}}{N(N+3)}. \end{aligned}$$

REFERENCES

- [1] B. Buraimo and R. Simmons, *Do sports fans really value uncertainty of outcome? Evidence from the English Premier League*, International Journal of Sport Finance **3** (2008), 146–155.
- [2] A. Dawson and P. Downward, *Measuring short-run uncertainty of outcome in sporting leagues*, Journal of Sports Economics **6** (2005), 303–313.
- [3] D. Forrest and R. Simmons, *Outcome uncertainty and attendance demand in sport: The case of English soccer*, Journal of the Royal Statistical Society, Series D **51** (2002), 229–241.
- [4] K. K. Haugen, *Point score systems and competitive balance in professional soccer*, Journal of Sports Economics **9** (1991), 191–210.
- [5] K. K. Haugen, *Always Change a Winning Team*, Tapir University Press, Trondheim, 2012.

- [6] K. K. Haugen and B. Guvåg. *Uncertainty of outcome and rule changes in European handball*, European Journal of Sport Studies, 2018, doi: 10.12863/ejssax6x1-2018x2.
- [7] K. K. Haugen and K. P. Heen, *The competitive evolution of European top football – signs of danger*, European Journal of Sport Studies, 2018, doi: 10.12863/ejssax6x1-2018x1.
- [8] W. C. Neale, *The peculiar economics of professional sports*, The Quarterly Journal of Economics **78** (1964), 1–14.
- [9] S. Rottenberg, *The baseball player's market*, Journal of Political Economy **64** (1956), 242–258.
- [10] A. S. Zimbalist, *Competitive balance in sports leagues: An introduction*, Journal of Sports Economics **3** (2002), 111–121.

Kjetil K. Haugen, Faculty of Business Administration and Social Sciences, Molde University College, Specialized University in Logistics, Britveien 2, 6410 Molde, Norway
e-mail: kjetil.haugen@himolde.no

