On bias in the ECF rating system

This note sets out the essential mathematics on the effect of bias in the ecf rating system.

Bias, positive or negative, exists when the grade, an estimate of player’s strength, differs from true strength. It can be seen that if all players play a large number of other graded players then overall their grades will tend to be unbiased. Nonetheless there will be bias for each grade, but at a universal level, assuming the initial calibration was unbiased; pluses and minuses tend to cancel out. Bias could be material distortion of grading for individuals when their grade is based on a small number of graded opponents.

More complete details of the ECF grading system are available on the ECF website. In its simplest case a player scores their opponents grade adjusted by their score (1,1,0) times 100 less 50. This result is averaged over the season and treated as an indicator of relative strength. In this note grades are taken as 1% of these numbers just to simplify annotation.

The graders use an iterative approach to get results of the following matrix equations. There is evidence that this approach gets the same answers. This justifies some loose matrix multiplication in the development of the ultimate equations.

It is worthwhile standing back and looking at why bias might exist[1]. A grade from one game can be broken down into a few components:

- Opponent’s true strength
- Bias in their grade
- Grading result (+50,0 or –50)
- Performance fluctuation by either party

The system implicitly assumes that the sum of the biases in opponent’s grades will tend to zero as the number of games increases. This assumption can be affected by matches (multiple games against another player). It could also be distorted by methodological biases, for example the 40 point rule, the treatment of juniors and overseas players or the recent recalibration exercise.

It also assumes that that the sum of opponent’s true strength plus grading result will tend to player’s true

[1] This section is based on a post on the ecforum by E Michael White

strength over a large number of games (deemed as 30. Furthermore the only material fluctuation identified in the system is for improving juniors.

This note starts by defining some terms, which are then used to construct a simplified model of the ecf grading algorithm and then finally looks at some statistical properties of that model.

1. Definitions

The mathematics ignore the 40 point rule and complications arising from treating categories below A as carrying forward older ratings. The final junior adjustment (+5 or +10) is ignored.

Define y as the number of players being rated. Currently, y is of the order of 10,000.

P^1 is an yx1 matrix of the final grades of those y players, multiplied by .01

D is an yxy diagonal matrix, where element d(i,i) is the number of games played by player i; d(i,j)=0 i<>j.

M is an yxy matrix, where element n(i,j) is the number of games played between players i and j.

N is an yxy matrix where element n(i,j) is the proportion of the graded games in the period by player i played against player j. This entry will invariably be 0 or 1/(games played by i). N is calculated as D^{-1}.M

W is a yxy matrix is the grading result adjustment, where element w(i,j)is [(S scores by i against j)-(0.5*games between i and j)]

S is an yxy matrix = D^{-1}.W

g is an yx1 matrix where element g(i) is the underlying strength of player i. This is what we are trying to estimate.

1 is a variable length column vector which in effect sums each row of the prior matrix.

Read p^{any} as performance, an estimate of g and let p^* be the reference grades (explained below)

Then the fundamental formula for ECF rating is:

$D.P^1 = M.p^* + W.1$

The general idea is that gradings are assumed not to change fast so that the reference grade is that of the previous year. However there are new entrants who have no prior grade and juniors who are being treated as new entrants, because many of them appear to be improving.

2. Model Development

To investigate further the matrix is partitioned between regulars, R - non juniors with a prior grade and others, U, who are treated differently.

Hence the fundamental formula is rearranged as

$P^1 = N.p^* + S.1$

then partitioned and sorted in a natural indexation as:

$$\begin{bmatrix} P^R \\ P^U \end{bmatrix} = \begin{bmatrix} N^{RR} & N^{RU} \\ N^{UR} & N^{UU} \end{bmatrix} \begin{bmatrix} p^0 \\ p^* \end{bmatrix} + \begin{bmatrix} S^{RR} & S^{RU} \\ S^{UR} & S^{NU} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where p^0 is the available prior year grade and p^* can be interpreted from a description of the process used by the graders as the solution to the following equation.

$$[P^*] = [N^{UR} \; N^{UU}] \begin{bmatrix} p^0 \\ p^* \end{bmatrix} + S^{UR} 1 + S^{UU} 1$$

The solution for p^* is:

$$P^* = [I - N^{UU}]^{-1} \{ N^{UR} p^0 + S^{UR} 1 + S^{UU} 1\} \qquad (1$$

Where I is the identity matrix.

This vector can be plugged into the partitioned matrix giving two matrix equations for the separate methods.

$$P^R = N^{RR} p^0 + N^{RU} p^* + S^{RR} 1 + S^{RU} 1 \qquad (2A$$

$$P^U = N^{UR} p^0 + N^{UU} p^* + S^{UR} 1 + S^{UU} 1 \qquad (2B$$

Substituting Equation 1 into equation 2A gives:

$$P^R = N^{RR} p^0 + S^{RR} 1 + N^{RU} [I - N^{UU}]^{-1} N^{UR} p^0 + N^{RU} [I - N^{UU}]^{-1} S^{UR} 1 + N^{RU} [I - N^{UU}]^{-1} S^{UU} 1 + S^{RU} 1$$

Equation 2B is resolved by plugging in equation 1 and re-arranging giving a solution:

$$P^U = \{I + N^{UU} [I- N^{UU}]^{-1}\}\{N^{UR} p^0 + S^{UR} 1 + S^{UU} 1\}$$

3. Statistical support

The underlying principle is that the expected scores should follow the relative performance standards. While this could be written game by game, the performance is over a year and the expectation here is taken in total as:

$$E(W.1) = Dg – Mg$$

$$E(S.1) = D^{-1}[Dg – Mg] = [I-N]g \qquad (3$$

Partitioning equation 3 gives:

$$E\begin{bmatrix} S^{RR} & S^{RU} \\ S^{UR} & S^{UU} \end{bmatrix}.1 = \left\{\begin{bmatrix} I^{RR} & 0 \\ 0 & I^{UU} \end{bmatrix} – \begin{bmatrix} N^{RR} & N^{RU} \\ N^{UR} & N^{UU} \end{bmatrix}\right\}.\begin{bmatrix} g^R \\ g^U \end{bmatrix}$$

Breaking this equation down gives:

$$E [S^{RR}.1] = I^{RR} g^R – N^{RR} g^R$$

$$E [S^{RU}.1] = -N^{RU} g^U$$

$$E [S^{UR}.1] = – N^{UR} g^R$$

$$E [S^{UU}.1] = I^{UU} g^U – N^{UU} g^U$$

Taking expectations of the fundamental equation gives, assuming [g] is fixed and known:

$E(P^1) = N.p^* + [I-N]g = N[p^* –g] + I.g$ and expanding:

$$E\begin{bmatrix} P^R \\ P^U \end{bmatrix} = \begin{bmatrix} N^{RR} & N^{RU} \\ N^{UR} & N^{UU} \end{bmatrix} \begin{bmatrix} p^0 - g^R \\ p^* - g^U \end{bmatrix} + \begin{bmatrix} I^{RR} & 0 \\ 0 & I^{UU} \end{bmatrix}\begin{bmatrix} g^R \\ g^U \end{bmatrix}$$

$$E[P^R] = N^{RR}[p^0 - g^R] + N^{RU} [p^* - g^U] + g^R \qquad (4A$$

$$E[P^U] = N^{UR}[p^0 - g^R] + N^{UU} [p^* - g^U] + g^U \qquad (4b$$

Starting with 4b:

re-arrange and substitute (1 into 4b

$$E[P^U - g^U] = N^{UR}[p^0 - g^R] + N^{UU}([I - N^{UU}]^{-1}\{N^{UR} p^0 + E[S^{UR} 1 + S^{UU} 1]\} - g^U)$$

$$= N^{UR}[p^0 - g^R] + N^{UU}([I - N^{UU}]^{-1}\{N^{UR} p^0 - N^{UR} g^R + I^{UU} g^U - N^{UU} g^U\} - g^U)$$

Hence

$$\boxed{E[P^U - g^U] = \{I + N^{UU}[I - N^{UU}]^{-1}\} N^{UR}[p^0 - g^R] \quad (5b}$$

Now return to 4a and continue in similar fashion:

$$E[P^R - g^R] = N^{RR}[p^0 - g^R] + N^{RU}[p^* - g^U]$$

$$= N^{RR}[p^0 - g^R] + N^{RU}([I - N^{UU}]^{-1}\{N^{UR} p^0 + E(S^{UR} 1 + S^{UU} 1)\} - g^U)$$

$$= N^{RR}[p^0 - g^R] + N^{RU}([I - N^{UU}]^{-1}\{N^{UR} p^0 - N^{UR} g^R + I^{UU} g^U - N^{UU} g^U\} - g^U)$$

$$= N^{RR}[p^0 - g^R] + N^{RU}[I^{UU} - N^{UU}]^{-1}\{I^{UU} g^U - N^{UU} g^U\} - g^U$$

Hence

$$\boxed{E[P^R - g^R] = \{N^{RR} + N^{RU}([I - N^{UU}]^{-1} N^{UR}\}[p^0 - g^R] \quad (5a}$$

If $E[P^0 - g^R]=0$ for a given [g] then $E[P^U - g^U]=0$. So if the system starts unbiased then in this simplified form it remains unbiased. However the methodology could be tripped out of an unbiased state. This could arise from mis-estimation of improvements, colour bias or percolation from isolated communities (e.g. scholastic chess).

These equations allow tests of the system. One test is then whether there is a natural correction. Another might look at the efficiency of the estimators. The issue about estimating [g] given it is not known introduces a feature that is considered elsewhere – that of observed stretch. These are all issues requiring further research.