Estimation of Golf Course Ratings from Player Scores

Peter Preston*
Statistical Consultant, Golf Australia, Melbourne, Australia
prestonp@bigpond.net.au

Golf handicaps are used to give players of different golf ability a roughly equal chance of winning a competition. However, for golfers on handicaps above zero, some courses are much harder to play than others. Players from hard courses are likely to have higher handicaps than players of the same ability from easy courses. When a player from a hard course goes to play at another easy course, that player's chance of winning may be greater because of the higher handicap. The United States Golf Association has developed a system called Slope to deal with this effect. Two separate Ratings, called Scratch and Bogey Ratings, are determined from physical measurements of a course. These Ratings are used in a formula which adjusts players' handicaps when they play at different courses. Considerable work is required to carry out the physical measurement and analysis to generate the Ratings for all golf courses. There are questions about how accurately these measured Ratings actually reflect the scores that players achieve on the courses. A methodology for estimating the Ratings from recorded golf scores is proposed here. The scores of players who play the course regularly, i.e. home players, cannot be used, because their handicaps have adjusted to their home course. Therefore the methodology principally uses the scores of visitors to a course to estimate the Ratings of that course. However where there are different tees in use at a course, as is very common, home players' scores can be used to estimate the difference in Ratings between the different tees. The methodology is being trialled in Australia. It requires a comprehensive database of golf scores and related information, such as the type of competition that was played. The formulae for estimation of golf course Ratings from scores are presented in this paper. Once the methodology is proven satisfactory, it has the potential to be implemented as an automated process which is much more economical and accurate than the current approach. It could reduce the need for physical measurement of golf courses to estimate course Ratings around the world.

Key Words: golf scores, scratch and bogey Ratings, golf course Slope, parameter estimation.

1. Handicaps

This paper deals only with stroke golf scores; equivalent formulae have been developed for stableford and par scores, and conversion is easy.

The principles that are developed in this paper are identical for men's and women's scores; the only difference is that some of the constants used vary by gender.

Golfers have handicaps calculated from their most recent scores. Essentially these handicaps are estimates of the golfer's underlying golf ability at this time. Handicaps are used to calculate: Net Score = Gross Score – handicap.
Net Score is commonly used to determine competition winners and place getters in handicapped competitions. The intention of this is for all players in a competition to have a roughly equal chance of winning or being a high place-getter.

Various different handicap algorithms are in use in different countries. The algorithms use players' recent "Played To" (PT) scores, where PT = S – SR.
Here S denotes stroke score, and SR denotes the course Scratch Rating (see Section 2).

For example, in Australia, handicap is calculated by an algorithm which makes use of the eight best (i.e. lowest) Played To scores from the last twenty rounds.
More precisely:
H = (Average of eight best) * 0.93 --- see www.golfaustralia.org.au.

It should be noted that the "Par" of a course, which features on golf score cards hole by hole, has no role to play in the calculation of handicaps, and is not mentioned further in this paper.

2. Course Ratings and Slope

Some golf courses are much harder to play than others. This is a differential effect in the sense that a scratch golfer (handicap 0) may, on average, take one more stroke on course Y than course X, whereas a bogey golfer (handicap about 20) may take four strokes more. Without some compensating mechanism, players from more difficult courses are more likely to have higher handicaps than players of the same ability from easier courses. Thus, when a player from a more difficult course goes to play at an easier course with no change in handicap, that player's chance of winning is likely to be greater because of the higher handicap.

The United States Golf Association has developed a system called Slope to deal with this effect, see --- www.usga.org. Scratch and Bogey Ratings (SR and BR) are determined from a range of physical measurements and other observable attributes of a course. The physical measurement process requires people to be trained and to contribute considerable time and effort. It involves a level of skill and judgement. These Ratings that are produced are predictors (in a statistical sense) of the scores that Scratch and Bogey golfers will achieve. They are used in a formula which adjusts players' handicaps when they play at different courses. This is intended to provide fairness to all players in competitions at all courses.

In concept, courses have "True" Ratings, TSR and TBR. These Ratings can never be known. Ratings produced from physical measurement are denoted by ME_SR and ME_BR. The "ME" in this terminology refers to "Measurement Estimate". In other words ME_SR and ME_BR are estimates produced by physical measurement of the True Ratings TSR and TBR.

This paper addresses the application of Slope methodology in a region with a number of golf courses, e.g. in a country such as Australia.

It is assumed that in the region there exists a "Standard Course". It has the average SR and the average BR of all courses in the region. These are denoted by ASR and ABR. A player whose handicap is obtained from a Standard Course has a "standard course" handicap. The symbol **h** is used to denote "Standard Course" handicap. A handicap obtained from a "non-standard" course is denoted by **H**. We define RD (Rating Difference) in general as Bogey Rating minus Scratch Rating: RD = BR – SR.
For True values of Ratings for any course TRD = TBR – TSR. Similarly for the Standard Course: ARD = ABR – ASR.

Suppose course X has Measured Estimated Ratings ME_BRX and ME_SRX, and Measured Estimated Rating Difference ME_RDX = (ME_BRX – ME_SRX). If a player at course X has a calculated handicap H, the corresponding calculated "Standard Course" handicap is:
h = H * ARD / ME_RDX.
This is referred to as "De-Sloping". It puts all handicaps onto the same baseline to produce a course - independent estimate of player ability.

When a player who has a calculated handicap HX from course X goes to play at course Y, the handicap is "Sloped" from course X to course Y by:

$$HY = HX * ME\_RDY / ME\_RDX,$$

i.e. it is Sloped by the ratio of the Ratings Differences. If these Measurement Estimates of the Ratings Differences of the two courses are in error from the True values, then the resulting Sloped handicap HY is in error. If the errors in individual Estimated Ratings are a couple of strokes in magnitude, it is possible for the errors in HY to be large and potentially cause significant problems in golf competitions. Visitors will play in competitions with handicaps that can be many strokes in error – if they are large positive errors the visitors will have a very large unfair advantage. If they are large negative errors, the visitors will have a large unfair disadvantage.

To the author's knowledge there has not been a validation study to assess the accuracy of the estimates ME_BR and ME_SR generated by the physical measurement process in countries such as Australia. In this context, accuracy means that these Ratings estimates predict players' scores accurately in a statistical sense. Evidence is accumulating that these Measurement Estimates can be substantially in error. To ensure that the application of Slope produces its desired effect in handicapped golf competitions, there is strong need to improve the accuracy of the Ratings estimates used for Slope.

In reality course difficulty varies day to day. The simplifying assumption that Ratings of a course are constant and equal to the long term average (over say a couple of years) is made in this paper. New processes to deal with daily variation have been developed recently in Australia, but are beyond the scope of this paper.

3. The Golf Score Equation

A player's score S is a random variable that depends on:

- the ability of the player, and
- the difficulty of the course.

Analysis of a large sample of golf scores indicates that the distribution of S is very close to Normal. There is a slight positive skewness, but it is not great enough to be an issue with the formulae presented here. Thus it is assumed that:

$$S \sim N(\mu, \sigma^2).$$

A player's ability is represented by True Handicap TH. Then:

$$\mu = TSR + TH + ND(TH) \tag{1}$$

Here TSR as the True Scratch Rating and TH is the player's True Handicap Sloped to this course. The final term ND(TH) refers to the "Normal Deduction", a term conceived by Michael Maher (Michael@statisticalsolutions.com.au).

Everyone who plays golf knows that a player's average score does not equal TSR + TH. Players do not, on average, play to their handicap. The Normal Deduction is the number of strokes by which the average score exceeds the Scratch Rating plus Handicap, TSR + TH.

Maher showed that the Normal Deduction is very strongly a linear function of handicap (TH), and he has estimated the parameters of the straight line for Australian golfers. The parameters are available for all six combinations of score type (stroke, stableford, par) and gender.

Then: ND(TH) = m* TH + b, where m and b are constants known by estimation from a very large sample of scores. Thus:

$$\mu \quad = \quad TSR + TH *(1 + m) + b \qquad (2)$$

Maher (2011) also showed that σ is strongly a linear function of TH, so

$$\sigma \quad = \quad m' * TH + b' \qquad (3)$$

Basically, as handicap increases, the Standard Deviation of score increases as a linear function of handicap.

4. Estimation Equations

In principle one can use equation (2) to estimate a course Scratch Rating. Suppose there are a sample of scores S_1,...S_n from a player on True Handicap TH; we compute the mean score $\overline{S}$, and substitute it for μ. Then

$$SE\_SR \quad = \quad \overline{S} \; - \; TH *(1 + m) \; - \; b$$

Here “SE_SR” refers to “<u>S</u>tatistical <u>E</u>stimate of <u>S</u>cratch <u>R</u>ating”. In practice a player’s True Handicap TH is not known. Each player has an estimated handicap H which is calculated from recent scores using the handicap algorithm, so this equation becomes:

$$SE\_SR \quad = \quad \overline{S} \; - \; H *(1 + m) \; - \; b \qquad (4)$$

It was noted in Section 1 that handicaps are calculated from Played To scores (PT = S – SR). If the SR value used in this calculation is in error from the True value, for example if SR = TSR – 2, then each PT value will be 2 strokes too high, and the resultant calculated handicap will be too high. In the case of the Australian handicap algorithm, the calculated handicap will be 2 * 0.93 = 1.86 strokes too high.

It follows that for players who play their home course most of the time (i.e. members), if the estimated Ratings for that course are in error, then the players’ handicaps H are in error by a similar amount. Therefore if the Ratings of the course are estimated from those players’ scores using equation (4), the Ratings estimates will also be in error by a similar amount.

In summary, members’ scores cannot be used to estimate primary Ratings (however in Section 6 it is shown how members’ scores can be used to estimate Ratings of secondary tees).

Now consider the situation with visitors’ scores. Visitors are defined as players who play at this course occasionally. Their handicaps are determined by Ratings of another course, or courses. If the Ratings of the courses which determine the visitors’ handicaps are in error, then the visitors’ handicaps are in error. Ratings are relative numbers. Scratch Ratings are relative to the ASR; roughly half of all courses have TSRs greater than ASR and half smaller. The TSR for the average course is equal to the ASR by definition. So if some courses have ME_SRs that are in error, e.g. too large, then the players whose handicaps are set by those courses will have handicaps that are too small, by roughly the same amount. Similarly, courses that have ME_SRs that are too small will have players whose handicaps are too high. So it is reasonable to assume that if there is a random sample of a large number of visitors to a course, and the visitors come from a large sample of courses, half of them will have handicaps too large, and half too small; the average handicap of visitors will be correct. That is, ***on average,*** visitors’ calculated handicaps will equal their True handicaps. So in essence, equation (4) can be used to estimate Ratings if the scores are visitors' scores from a large sample of visitors who play at a large sample of different courses, and it can be assumed that ***on average*** across all visitors' scores **H** is correct.

The process of estimation of the Ratings TBR and TSR of a course (call it Course X) starts with a sample of visitors' scores $S_1,...S_n$ at Course X. The estimates that are produced are called SE_BR and SE_SR.

Consider one visitor score S_i. That visitor's Standard Course handicap h_i is calculated from all scores by that visitor prior to round i, with any scores at Course X being ignored. This eliminates any impact the Estimated Ratings of Course X could have had on the calculated handicap of this visitor.

Each Played To score used in the calculation of h_i has been de-Sloped to the Standard Course. This de-Sloping makes use of the Measured Estimated Ratings of all the other courses at which the visitor played, but does not involve the estimated Ratings of Course X. The process is repeated for every score, $S_1,...S_n$. The sample of points $\{S_i, h_i\}$ i.e. {scores, calculated standard handicaps} is used to produce a (weighted) regression equation:

$$\text{MeanScore}(h) = A * h + B \qquad (5)$$

Equation (3) indicated that the score data is heteroscedastic. In fact it is an unusual regression situation because the variance $(m^l * h + b^l)^2$ is known quite precisely. A weighted regression is performed with weights inversely proportional to the variance of each data point.

Equation (2) is in terms of TH, which is true Sloped Handicap at Course X. Then
TH = h * (TBR – TSR) / ARD,
so (2) can be re-written:

$$\mu = TSR + h * (1 + m) * (TBR - TSR) / ARD + b \qquad (6)$$

Equation (5) gives the regression line mean score as a function of standard handicap **h**, and equation (6) gives it as function of **h** and the parameters to be estimated, TSR and TBR. Substituting the Estimated Parameters SE_SR and S_BR for TSR and TBR into (6), and solving the simultaneous equations gives:

$$SE\_SR = B - b \qquad (7)$$
$$SE\_BR = SE\_SR + ARD * A / (1 + m) \qquad (8)$$

These equations can be interpreted easily. For a Scratch golfer the value h = 0 is used, and the estimate of TSR is the MeanScore of a scratch golfer read from the regression line, i.e. B, minus the Normal Deduction for a scratch golfer, i.e. b.

A Bogey golfer on handicap h = ARD has a MeanScore read from the regression line of A * ARD + B. Substituting h = ARD, and SE_SR and SE_BR, into the right hand side of (6) gives:

SE_SR + (1 + m) * (SE_BR – SE_SR) + b. Thus:
A * ARD + B = B + (1 + m) * (SE_BR – SE_SR), which again gives (8).

Results of Ratings Estimates calculated from these equations using real golf scores are expected to be presented in a future paper. However it is noted that care must be taken to ensure the data is well conditioned and representative. For example, the author recommends that no more than 5 visitors' scores from any one person be used, and it should be ensured that the sample of courses from which the scores used to calculate the visitors' handicaps come be diverse and representative.

5. Standard Errors of Ratings Estimates from Visitors' Scores

For weighted regression equations with heteroscedastic data:

$\text{Var}(\boldsymbol{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X} \;\; \boldsymbol{\psi} \;\; \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$, where: $\boldsymbol{\beta}$ is the vector of regression parameters, in this case the 2 x 1 vector $[B, A]^T$. **X** is the n x 2 matrix of independent variables. Here we use h for handicap, in place of X as the independent variable. Then the first column of **X** is$[1,1...1]^T$; and the second column is $[h_1, \ldots h_n]^T$. $\boldsymbol{\psi}$ is the variance – covariance matrix of residuals, in this case a diagonal matrix with $\psi_{ii} = \sigma_i^2 = (m^{I} * h_i + b^{I})^2$; see equation (3). In (7) and (8) the only parameters which have been estimated by regression are A and B, so that the ***Standard Errors*** of the Ratings Estimates are:

SE {SE_SR} = √ Var(B), and

SE {SE_BR} = √{Var(B) + 2*ARD*(1+m) * Cov(B,A) + $[\text{ARD}*(1+m)]^2$ * Var (A)}.

6. Estimation of Secondary Tee Ratings from Members' Scores

Many golf courses have multiple tees. The men's tee from which there is the greatest number of recorded visitors' scores is defined as the men's primary tee, and similarly for the women's primary tee. The estimates SE_SR and SE_BR of SR and BR of the primary tees are calculated from visitors' scores using (7) and (8). Other tees are defined to be secondary tees. For the purpose of estimating the Ratings of the secondary tees, SE_SR and SE_BR are assumed to be the True values of the Ratings of the primary tees, and the Ratings of the secondary tees are estimated from members' scores at those tees. For each member's score used in the estimation equation, the member's primary tee course handicap is re-calculated using:

- only scores at the primary tee,
- SE_SR at the primary tee as the value of SR in the handicap calculation,

  and
- each Played To score de-Sloped using the ratio ARD/ (SE_BR – SE_SR).

The Ratings estimates of the secondary tees are then calculated from the sample of points $\{S_i, h_i\}$ of members' scores at the secondary tee and corresponding primary tee course handicaps using (7) and (8).

The advantage of this approach is that usually there are many more members' scores at the secondary tees than there are visitors' scores, so the corresponding Ratings estimates will be more accurate. A variation on this that can be used is to combine members' scores and visitors' scores into a larger sample.

7. Conclusions

Formulae for estimation of golf course Ratings from recorded golf scores have been presented. The formulae can be implemented in a centralized golf score database for a country or region. The process is potentially more accurate and much more economical than estimating Ratings from physical measurement; with the proviso that the score data must be of high integrity, e.g. the scores were correctly and reliably obtained and recorded, and the course played was the same as the course being rated. It has the potential to make the application of Slope more effective and thereby achieve a higher degree of fairness to all golfers in golf competitions which include visiting players.

8. Acknowledgements

The author wishes to acknowledge the strong support and encouragement given by Golf Australia for this work, and excellent collaboration with Michael Maher in the development, testing, and evaluation of the concepts presented here.