**1. Introduction**

The game is golf is a test of a variety of skills – driving off the tee, approach shots from the fairway or rough, and putting to name a few. Precise estimation of the skill of the players in the various aspects of the game is useful for a variety of reasons. With accurate estimations of how players’ skill sets compare, players and coaches can create data-driven training plans and fans watching the game can gain a greater understanding of the strengths and weaknesses of their favorite players.

This paper improves what is currently being done to estimate the skill levels of the players on the PGA Tour. This work would not have been possible without detailed shot-level data that the PGA Tour started collecting in 2003 using their ShotLink™ system. The availability of these data has opened up the possibilities towards understanding the professional game. Up until detailed shot-level data was collected, it was impossible to quantify how the distinct skills determine golfers’ results.

This work also owes its foundation to the work done by Mark Broadie of Columbia University. His work in developing the Strokes Gained concept (explained in Section 2) has advanced everyone’s understanding of the game by being the first to really quantify individual skill sets of the players on the PGA Tour. His contributions and the work of others in this area are summarized in Section 2 of this paper.

Quantifying players’ skills is a pursuit of estimating latent variables. Competitive play is not setup in a way that makes this estimation simple. It is far from a scientific experiment where players are told to take multiple attempts from precise locations under controlled circumstances. In golf, players take around 72 shots per round but every shot is unique. A slight change of angle can make a shot entirely different. The quality of the lie can make two shots taken very close to one another very different. Weather conditions can vary from the morning to afternoon.

The challenges involved in this modeling problem will be detailed first. Then a novel approach will be given. This approach is then backed with evidence that demonstrates its success.

**2. Strokes Gained**

Before the detailed shot-level data and the Strokes Gained concept, statistics used to quantify specific skills included Driving Distance, Fairways Hit, and Greens In Regulation (GIR) to name a few. To illustrate the ambiguity that results from these statistics, take GIR as an example. GIR is the count of the number of holes on which a golfer reaches the green in two strokes less than the par value of the hole or fewer. GIR attempts to quantify a golfer’s skill with his or her approach shots. However, if two players start from the same position in the fairway and one hits it on the green 80 feet away and the other hits it to the fringe 18 feet away, the player who hit it on the green will be credited with a GIR while the other player will not, despite having left his ball in (arguably) a less desirable position.

This example illustrates the need to quantify the “desirability” of a particular location on a particular course on a particular day, or equivalently the difficulty of playing a shot from a particular location. It also motivates quantifying the quality of a particular shot by taking the difficulty of the starting location and subtracting the difficulty of the finishing location. This is the idea developed by Mark Broadie and is named the Strokes Gained Statistic.

To continue with the previous example, if the two golfers started from the fairway where it takes an average golfer 3.3 strokes (which tends to correspond to about 225 yards on tour), and we know how difficult it is for the average golfer from the locations where the two golfers’ balls ended up, we can quantify the quality of both players’ shots. From 80 feet the average PGA Tour golfer takes about 2.3 strokes to get the ball in the hole on average, while from 18 feet away on the fringe the average tour player takes about 1.9 strokes to get it in on average. Following the convention established in Broadie (2008), the Strokes Gained Statistic is then calculated using the following equation:

.

To conclude the example, the player whose ball ended up on the green had a shot quality of 0 (3.3 – 2.3 – 1), while the player whose ball ended up on the fringe had a shot quality of 0.4 (3.3 – 1.9 – 1). A positive shot quality corresponds with a shot that was better than the average player would have done and a negative shot quality means a shot that was worse than the average player would have done.

*2.1 Assumptions of Strokes Gained System*

Before continuing towards making a model of how difficult a given shot is, it is useful to think about the assumptions of the Strokes Gained framework. The first assumption is that we can estimate with reasonable accuracy how difficult a shot is. This is actually quite a challenge and there are potential pitfalls in doing this, which will be discussed shortly.

The second assumption is more fundamental. What does it mean to quantify the difficulty of a given shot? In Broadie (2008) this is defined as the average number of strokes taken from a given location by an average player. There is a subtle assumption in this method – that the desirability of a given location is the same for all players. This is generally a safe assumption because it’s mostly true; the desirability of different locations is very similar for all players. However, it’s useful to acknowledge that this method is a simplification of how the game is actually played. A consequence of this simplification is that the possibility that a player acts strategically is ignored. For example, a player could be faced with an option to lay-up on a par five, or try to hit it on the green, which is surrounded by bunkers. If this player is an excellent bunker player, this will certainly factor into his decision about whether or not to go for it. However, post-hoc evaluation of the quality of this players’ shot will take into account the desirability of the location he ends up in as measured by the theoretical performance of an average golfer from that location and thus will not correctly account for the strategic thinking that was involved in playing the shot.

This work will focus on coming to terms with the first assumption. The second assumption is more complex and will be left for another contributor.