PROGRAM

September 24, 2011

Harvard University
Science Center, Lecture Hall C
1 Oxford Street
Cambridge, Massachusetts 02138

Symposium Co-Organizers:

Mark E. Glickman, Department of Health Policy and Management, Boston University School of Public Health

Jason W. Rosenfeld, Department of Statistics, Harvard University

Scott R. Evans, Department of Biostatistics, Harvard University School of Public Health

Sponsors:

- Boston Chapter of the American Statistical Association (`http://www.amstat.org/chapters/boston/`)
- Section on Statistics in Sports of the American Statistical Association (`http://www.amstat.org/sections/sis/`)
- Harvard University Department of Statistics (`http://www.stat.harvard.edu/`)
- StatDNA (`http://www.statdna.com/`)
- SmartOdds (`http://www.smartodds.co.uk/`)
- Sports Data Hub (`http://www.sportsdatahub.com/`)
- Salford Systems (`http://salford-systems.com/`)
- ESPN (`http://www.espn.com/`)

Acknowledgments: We wish to thank everyone who helped to make NESSIS possible. Special thanks go in particular to Tom Lane for creating and maintaining the NESSIS web site and sending out BCASA announcements. We would also like to thank Rick Cleary, Kevin Rader, Andrew Thomas, Rick Peterson, Rosanne Desmone, Sowmya Rao, Mike Zarren, Carl Morris, Betsey Cogswell, Ellen Weene, Dale Rinkel, Rachel Breinin, Columbia Pictures, Iram Farooq, Stefan Cheplick, and Huichao Chen for their parts in helping with the symposium.

2011 New England Symposium on Statistics in Sports

September 24, 2011

All talks take place in the Harvard University Science Center, Lecture Hall C.
All breaks will be in the foyer area outside the lecture hall.

9:15am – 9:30am:	Welcome
9:30am – 10:00am:	James Piette, Krossover Intelligence Inc. *"The value of teamwork: Estimating individual performance via a statistical network algorithm"*
10:00am – 10:30am:	Samuel Ventura and Andrew Thomas, Carnegie Mellon University *"Improving NHL player ability ratings with hazard function models for goal scoring and prevention"*
10:30am – 11:00am:	**StatDNA contest winner presentation:** Sarah Rudd, On Football Research and Consulting *"A framework for tactical analysis and individual offensive production assessment in soccer using Markov chains"*
11:00am – 11:30am:	Break (coffee and tea)
11:30am – 12:00pm:	Adrian Schembri, RMIT University *"Comparing market efficiency with traditional and non-traditional ratings systems in ATP tennis"*
12:00pm – 12:30pm:	David J. Marcus, Ratings Central *"Ratings Central: Accurate, automated, Bayesian table tennis ratings for clubs, leagues, tournaments and organizations"*
12:30pm – 1:00pm:	Todd L. Graves, Los Alamos National Laboratory *"RUSH: Ratings using score histories"*
1:00pm – 2:00pm:	Lunch break (lunch provided)

2:00pm – 2:30pm: J. P. Ricciardi, New York Mets
Scouting analytics: Use of statistical information in MLB personnel decisions.

2:30pm – 3:00pm: Sheldon H. Jacobson, University of Illinois
"Seed distributions for NCAA men's basketball tournament: Why it may not matter who plays whom"

3:00pm – 3:30pm: Chris Stride, University of Sheffield
"Cheating in football: Team culture, player behavior, or question of circumstance?"

3:30pm – 5:00pm: Poster Session, with snacks and beverages

5:00pm – 6:30pm: Panel Discussion – *"Basketball analytics: Strategizing in the NBA front office."*

Moderator: Zach Lowe – NBA writer for `SI.com`, the Sports Illustrated web site

Panelist: Kenny Atkinson – Assistant Coach, New York Knicks
Panelist: Roland Beech – Director of Basketball Analytics, Dallas Mavericks
Panelist: Sam Hinkie – Executive VP of Basketball Operations, Houston Rockets

7:00pm – 10:00pm: Post-NESSIS get-together at

Tavern in the Square, Porter Square
1815 Massachusetts Ave., Cambridge, MA 02140
`http://taverninthesquare.com/tavporter/`

Oral Presentation Abstracts

RUSH: RATINGS USING SCORE HISTORIES

Todd L. Graves†[1], Earl Lawrence[1], Kary Myers[1], C. Shane Reese[2]

Statistical Sciences Group, Los Alamos National Laboratory[1]*, Department of Statistics, Brigham Young University*[2]

† E-mail: *tgraves@lanl.gov*

We present a model for rating teams that takes into account the score of each game at every point in time. Unlike pure margin-of-victory models, our methodology does not reward unnecessarily lopsided victories. This approach is of special interest for the NCAA FBS national championship, which is contested by two teams chosen after a regular season in which most teams never play each other. Ranking systems based on win-loss results, such as those used by the BCS, can have difficulty evaluating undefeated teams with weak schedules because of the limitations of these data. Alternative rating systems based on margin of victory can be strongly influenced by extreme results and may incentivize poor sportsmanship. Instead, our model considers the entire history of within-game scoring plays to infer the relative abilities of all teams. RUSH treats the score difference throughout each game (visitor minus home) as a stochastic process in order to model the evolution of the visiting team's probability of winning the game. Thus meaningless scoring plays, those that occur when the probability of winning is near 0 or 1, have little effect on the teams' ratings.

SEED DISTRIBUTIONS FOR NCAA MEN'S BASKETBALL TOURNAMENT: WHY IT MAY NOT MATTER WHO PLAYS WHOM

Sheldon H. Jacobson†[1], Alexander G. Nikolaev[2], Douglas M. King[1], Adrian J. Lee[3]

University of Illinois[1]*, University of Buffalo (SUNY)*[2]*, Central Illinois Technology and Education Research Institute*[3]

† E-mail: *shj@illinois.edu*

Bracketology, the art of successfully picking all the winners in the National Collegiate Athletic Association's (NCAA) annual men's Division I college basketball championship tournament, has become a favorite national activity. In spite of the challenges and uncertainty

faced in this endeavor, patterns exist in how the seeds appear in each round, particularly in the later rounds. This paper statistically analyzes tournaments from 1985 to 2011, finding that the distribution of seeds that win in the rounds beyond the Sweet Sixteen can be modeled as a truncated geometric random variable. This model allows one to consider any set of seeds in each tournament round and compute the probability that these seeds would win in that round; this methodology can evaluate the likelihood of each seed combination in each tournament round, based on past tournament history. Finally, each tournament from 1985 through 2011 is analyzed using this model to assess its likelihood and measure the probability of its occurrence. The resulting model was implemented in the website, `bracketodds.cs.illinois.edu`, prior to the 2011 NCAA Tournament. The web site attracted over 8,000 visitors during the period around the tournament. The key implication of the model is that the teams in the tournament and who they play is far less important than where they are seeded, to determine their likelihood of advancing deep into the tournament.

RATINGS CENTRAL: ACCURATE, AUTOMATED, BAYESIAN TABLE TENNIS RATINGS FOR CLUBS, LEAGUES, TOURNAMENTS AND ORGANIZATIONS

David J. Marcus[†1]

Ratings Central[1]

[†] E-mail: *davidmarcus@alum.mit.edu*

Ratings Central (`www.ratingscentral.com`) has been providing ratings for table tennis players in the U.S. and other countries since 2004, and has been adopted as the official rating/ranking system for several organizations including the Austrian Table Tennis Association. The system contains 32,000 players and 690,000 matches. Players range from five-year-olds to world champions. Most events are submitted by the event director, but we also extract results for the ITTF (International Table Tennis Federation) Pro Tour and Junior Circuit from the Web. The system is fully automated: an event director submits an event and the results appear on the website within minutes. If a correction is submitted, all affected events are automatically reprocessed. The system uses a Bayesian model with event directors providing priors for unrated players, specifying them either for a group (e.g., league division, tournament, age, gender) or individually. To make the Bayesian model computationally tractable, the graph of matches in an event (nodes for players, edges for matches) is modified for each player while still retaining most of the information on how a player's opponents did in the event. The system occasionally has trouble dealing with special subpopulations or players whose playing strength jumps. Although players and event directors sometimes have various misconceptions, in general they are very satisfied and agree that the

system is superior and more accurate than rating systems provided by USA Table Tennis or other table tennis organizations.

THE VALUE OF TEAMWORK: ESTIMATING INDIVIDUAL PERFORMANCE VIA A STATISTICAL NETWORK ALGORITHM

James Piette[†1], Lisa Pham[2]

Krossover Intelligence Inc.[1]*, Dept of Biomedical Engineering, Boston University*[2]

[†] E-mail: *jamesmpiette@gmail.com*

The major difficulty in evaluating individual player performance in any team sport is adjusting for interaction effects by teammates. With the advent of play-by-play data, the plus-minus statistic was created to address this issue. While variations on this statistic (e.g. adjust plus-minus [Rosenbaum 2004] and Defensive Efficiency Ratio [James 2001]) do correct for some existing confounders, they struggle to gauge two aspects: the importance of a player's contribution to his units or squads, and whether that contribution came as unexpected (i.e. over- or under-performed) as defined by a statistical model. Piette et al [2011] adapt a network-based algorithm [Pham 2011] to estimate centrality scores and their corresponding statistical significances. Players are treated as nodes in a network of every available player, where an edge exists between two players in they played in the same unit. The edges in this network are assigned weights corresponding to a function of the two players' performance during the time they played together. We determine the statistical contribution of a player in this network by comparing their centrality score in the actual network to a reference distribution generated from bootstrapping the edge weights on the original network. Building on this work, we improve on the previous modeling procedure used for basketball and apply the model to two new sports: baseball (fielding) and hockey.

A FRAMEWORK FOR TACTICAL ANALYSIS AND INDIVIDUAL OFFENSIVE PRODUCTION ASSESSMENT IN SOCCER USING MARKOV CHAINS

Sarah Rudd[†1]

On Football Research and Consulting[1]

[†] E-mail: *srudd@onfooty.com*

Markov Chains are an effective way to model transitions between states. Assuming that the current state is independent from the previous state, Markov Chains can be used to model the set of state transitions that make up a possession in soccer. The transitions are used to determine the probability a possession ends in one of two final states; scoring a goal or relinquishing possession to the opposing team. Once the final probabilities are known for each state, they can be used to determine game situations from which goals are more likely to develop, team strengths and weaknesses and metrics for assessing the offensive contributions of players.

Using this framework on the sample data set, we found that teams are more likely to score from taking long corners than short corners, with the notable exception of Tottenham Hotspur who excel at short corners. The top 3 teams most likely to score from a long corner are: Arsenal, Newcastle and Stoke. The top 3 teams most likely to concede from a long corner are: Everton, Arsenal and Newcastle. The framework can also be used to look at various game situations like building from the back, counter-attacks, free kicks, and entries into the final third, for example.

Additionally the transition probabilities can be used to determine which individuals are best at receiving the ball in situations with a high probability of scoring and which individuals are best at moving the ball to an improved state with a higher probability of scoring than their current state. The top 3 players for increasing the probability of scoring are Tim Cahill, Yaya Toure and Cesc Fabregas. The 3 most wasteful players who decrease their teams probability of scoring the most are Darren Bent, Peter Odemwingie and Gael Clichy. The top 3 players who receive the ball in the most advantageous states are Dimitar Berbatov, Nile Ranger and Benjani Mwauruwari.

COMPARING MARKET EFFICIENCY WITH TRADITIONAL AND NON-TRADITIONAL RATINGS SYSTEMS IN ATP TENNIS

Adrian Schembri†[1], Anthony Bedford[1]

School of Mathematical & Geospatial Sciences, RMIT University, Melbourne, VIC, Australia[1]

† E-mail: *adrian.schembri@rmit.edu.au*

Numerous factors can influence the ability to predict match outcomes in ATP tennis. These include surface factors, travel fatigue, injury status, and the current form of both players. Over the past decade, international men's tennis has endured a period of relative rankings stability, with a small number of players occupying the top rank and seed positions. Despite this, considerable variation has been evident in tennis betting markets during this period, with differing levels of volatility evident across various markets. In this paper, we evaluate the efficiency of several markets over the past ten years, with comparisons made between these markets and ratings systems such as Elo and SPARKS. Initially proposed by Bedford and Clarke (2000), the SPARKS model provides a unique alternative to the Elo system by rewarding players for winning games within a tennis match, with bonus ratings allocated for winning sets. This methodology is particularly salient in tennis given that consistently winning even a single set against a much higher ranked player is often indicative of an emerging player. By evaluating the efficiency of rating systems and comparing this to market odds, inefficiencies in the market can be readily identified. In addressing this, we investigate where inefficiencies in the market lie, the characteristics of matches that are associated with these inefficiencies (e.g., differences in player rankings, surface, tournament factors), and whether these inefficiencies are consistent across seasons or vary from year to year.

CHEATING IN FOOTBALL; TEAM CULTURE, PLAYER BEHAVIOUR OR QUESTION OF CIRCUMSTANCE?

Chris Stride†[1], Malcolm Patterson[1], Ffion Thomas[1]

University of Sheffield, UK[1]

† E-mail: *c.b.stride@sheffield.ac.uk*

We examined cheating, categorised in a sporting scenario by Loland (2005), in the context of association football, extending his typology by grouping 17 offences into 4 sub-types: Simulation, Classic Cheating, Professional Fouls and Calculated Dissent.

We hypothesised that variation in cheating would exist between players, teams and matches. Between-player variation would exist due to differences in mental and physical drivers, such as moral functioning, ability, experience, playing role. Between-match variation would be due to differences in match importance, competitiveness, and law enforcement. Between-team variation would be due to varying ability, expectation, style, "team climate," and national culture.

We collected data from the 2010 Football World Cup. Two coders watched every match, recording each cheating incident. Data was also collected for players (e.g. experience, position), teams (ability, manager's background, Hofstede's national cultural dimension scores; Hofstede, 2003), and matches (e.g. importance, referee).

For sub-types of cheating, multilevel generalised linear models were used to partition variation in the number of cheating events by player, team, and match, and test effects of player, team and match variables in explaining this variation. 412 cheating incidents were recorded, primarily Professional Fouls (275) and Simulation (83). For Professional Fouls, variation was found between matches and between players. This was predicted by role (defenders commit more), experience (experienced players commit less) and importance of match. For Simulation, variation was found between players and between teams. This was predicted by playing position and cultural dimensions, the latter mirroring Franke & Nadler's (2008) findings re: ethical behaviour within business organisations.

IMPROVING NHL PLAYER ABILITY RATINGS WITH HAZARD FUNCTION MODELS FOR GOAL SCORING AND PREVENTION

Samuel Ventura[1], Stephen Ma[1], Andrew C. Thomas[†1]

Department of Statistics, Carnegie Mellon University[1]

[†] E-mail: *act@acthomas.ca*

The public availability of data on NHL games now includes information on which players were on the ice for each team at any point during the game, and can naturally be divided into "shifts." In terms of goals scored during these shifts, the data are considerably sparse; the vast majority of shifts, on the order of one minute long, contain a goal by either team, making standard regression tools difficult to implement in assessing the contribution of each player to any goal scored for either team, including the weighting problem provided by shifts of different lengths. Additionally, the very nature of shift changes, in which player changes mean that the chances of scoring a goal are reduced, negate the use of a standard Poisson process. We turn these disadvantages into strengths by modelling the scoring rates

for each team as semi-Markov processes, whose hazard functions depend on the players on the ice for each team. We show that this method not only produces more stable estimates of player ability (including goaltending skill), it allows us to identify possible interaction effects between pairs of players who may be better suited to playing with each other.

Poster Presentation Abstracts

PARSING THE RELATIONSHIP BETWEEN BASERUNNING AND BATTING ABILITIES WITHIN LINEUPS

Ben Baumer[12], James Piette[†3], Brad Null[45]

CUNY Graduate Center[1]*, New York Mets*[2]*, Krossover Intelligence Inc.*[3]*, AOL Research*[4]*, www.baseballcalculus.com*[5]

[†] E-mail: *jamesmpiette@gmail.com*

A baseball team's offensive prowess is a function of two types of abilities: batting and baserunning. While research exists on each of these aspects in isolation, there has been little study on their interaction. In this paper, we look to gain new insight into this fundamental relationship. Building on the work of Baumer [2010], Terlecky [2010] and Null [2009], we assess representative batting and baserunning abilities for particular players within Major League Baseball. Then, using the simulation engine developed in Baumer [2009], we estimate the expected offensive performance of lineups composed of these players, as well as the gradient of this performance with respect to particular baserunning abilities and decisions. Our approach allows us to generate new insights into what are the most valuable baserunning abilities, what are the most impactful baserunning decisions, under what circumstances players in particular lineups should steal or attempt to "take the extra base," and how these conditions vary among different types of lineups.

METHODS TO REVERSE ENGINEER SPORTS RATING FORMULAS APPLIED TO THE SAGARIN RATINGS

Raymond Cheong[†]

Department of Biomedical Engineering, Johns Hopkins University

[†] E-mail: *rcheong@gmail.com*

The secrecy surrounding the mathematical formulas underlying well-known sports ratings prevents them from being independently reproduced, verified, and improved. Deducing the formulas has generally been considered to be a daunting task and there is currently a dearth of available algorithms for reverse engineering sports ratings. To fill this gap, I developed methods to decrypt sports ratings by a structured comparison of ratings values between successive (e.g. daily or weekly) updates, and I demonstrate that Jeff Sagarin's (unbiased) ratings are susceptible to these methods. In particular, I find strong evidence that the "PREDICTOR" ratings is based on weighted linear regression using uncapped score margins. "ELO CHESS" is consistent with the Elo rating equation with K factor of 1 and an exponential function for the expected game outcome. Also, the "RATING" value is a deterministic function of the "PREDICTOR" and "ELO CHESS" ratings, approximately the average of these two ratings adjusted by the degree to which they differ. In addition to presenting the first detailed examination of the mathematical formulas underlying the Sagarin ratings, I outline the steps needed to better understand the choice of parameters used in the formulas in order to generate a general, exact reconstruction. Finally, it may be possible to use similar methods to reverse engineer other gold standard ratings and enable them to be scientifically studied in a detailed manner.

STATISTICAL THEMES IN MODERN SPORTS METRICS

Rick Cleary[†1]

Dept of Mathematical Statistics, Bentley University[1]

[†] E-mail: *rcleary@bentley.edu*

As sports analytics have become increasingly sophisticated certain trends have emerged as common among many sports. A simple example is the growing use of rate measures rather than counts in player evaluation, such as RBI rates in baseball and percentages of available rebounds in basketball. We examine this and several other approaches (plus/minus, player value ratings, etc.) that are being employed in a variety of sports. We note the

similarities and differences in how these measures are (or should be) interpreted in the context of particular events. We also consider some statistical properties of these new classes of metrics, pointing out the cases in which they provide improvements over traditional measures in terms of robustness, reliability and usefulness in predicting performance. While our most detailed examples come from baseball and basketball we extend the idea to several other team and individual sports.

OPTIMAL FORECASTING OF REBOUNDING IN THE NBA

Jeremias Engelmann†[1]

Karlsruhe Institute of Technology, Germany[1]

† E-mail: *jeremias.engelmann@googlemail.com*

We compare the accuracy of three methods to predict the rebounding team for missed shots in NBA games. The first method uses individual rebounding numbers. Different techniques for splitting team rebounds between players and for handling rarely seen players are tested. In the second method, called "Adjusted Rebounding," ridge regression is used to find the player values in the context of rebounding. This method measures individual influence on team rebounding while adjusting for the rebounding skill of teammates and opponents. The third method also uses ridge regression, but has different penalizing parameters for the players according to their primary position on the court. Combinations of the methods are tested to find the optimal forecasting method. The results show that Adjusted Rebounding significantly outperforms individual rebounding numbers in forecasting. A combination of the first and third method outperforms each possible combination of the three methods.

CLUSTERING PERFORMANCE CURVES

Gilbert W. Fellingham†[1], Jared D. Fisher[1], Kristopher R. Young[1]

Department of Statistics, Bringham Young University[1]

† E-mail: *gwf@byu.edu*

Modeling changes across time is an important methodology for detecting individual or group differences in a variety of settings. Certainly, the performance curves of athletes offer insight into past contributions and perhaps into future expectations. We model the performance through time of major league baseball players who are in the top 200 all time for either batting

average or number of home runs hit using non parametric Bayesian methods. The data are assumed to be from a binomial distribution with either hits or home runs in a season as successes, and at bats as number of trials. The logistic transform of the probability of success is modeled as a linear function of orthonormal polynomials up to the fourth power based on the athlete's age. A Dirichlet process prior is used for the coefficients of the polynomials with normal centering distributions and a gamma prior for the DP precision parameter alpha. The clustering induced by the DP formulation is assessed using the method proposed by Dahl (2006), and a final clustering in produced based on minimum sums of squares. One interesting aspect of the result is that Babe Ruth, Mark McGwire and Barry Bonds in the home run analysis, and Ty Cobb and Joe Jackson in the batting average analysis are each in a single individual group. The induced clusters offer a unique insight in the discussion about Hall of Fame candidates.

A MARKOV MODEL OF FOOTBALL: USING STOCHASTIC PROCESSES TO MODEL A FOOTBALL DRIVE

Keith A. Goldner[†1]

Dept of Mathematics, Northwestern Univ[1]

[†] E-mail: *keithgoldner@u.northwestern.edu*

A team is backed into a 4th-and-26 from their own 25, down 3 points. What are the odds that drive ends in a field goal? In the 2003 playoffs, Donovan McNabb and the Eagles scoffed at such a probability as they converted and ultimately kicked a field goal to send the game into overtime. This study creates a mathematical model of a football drive that can calculate such probabilities, labeling down, distance, and yardline into states in an absorbing Markov chain. The Markov model provides a basic framework for evaluating play in football. With all the details of the model.absorption probabilities, expected time until absorption, expected points.we gain a much greater situational understanding for in-game analysis.

A STEP-BY-STEP INTRODUCTION TO DATA MINING FOR SPORTS ANALYSIS: A REAL-WORLD BASEBALL EXAMPLE

Mikhail Golovnya[1]

Salford Systems[1]

† E-mail: *nancyb@salford-systems.com*

Attendees will be provided with a step-by-step real-world baseball example illustrating the use of data mining as a powerful extension to classical statistical analysis. Data mining tools are able to find patterns in data and variable interactions that would be difficult, if not impossible, to detect using standard approaches. Attendees will gain new insights as to how data mining can quickly and easily find new indicators to measure performance and investigate any combinations of variables in your dataset. In the baseball example, data mining provides previously-unknown information related to predictors of true player/team potential, performance and success.

While just beginning to be of interest to the sports community, data mining has been successfully used for fraud detection at major credit card companies, drug safety research in the pharmaceutical industry, adserving for online advertising, promotional advertising optimization in the retail arena, and more. Data mining offers better insights as compared to standard statistical techniques when interactions (context dependant effects) are important, the data is dirty, or there are many missing values. Another strength of data mining is the ability to develop models rapidly and largely automatically. The methodology used for illustrative purposes in this presentation is based on algorithms developed by statistics and data mining leaders at Stanford University and the University of California at Berkeley: Dr. Jerome Friedman and the late Dr. Leo Breiman. All attendees will be given access for ninety days to our data mining tools so that they can try data mining on their own sports datasets.

DEFENSIVE REGRESSION ANALYSIS: ESTIMATING RUNS "SAVED" BY INDIVIDUAL PITCHERS AND FIELDERS THROUGHOUT HISTORY USING OPEN-SOURCE DATA

Michael A. Humphreys†[1]

Ernst & Young LLP[1]

† E-mail: *michael_humphreys@aya.yale.edu*

No rigorously derived statistical model based on publicly-available data has significantly improved baseball player valuation since Markov chain principles introduced by Lindsey (1963) culminated in Palmer's (1978) linear equations for estimating runs 'created' on offense above or below the league average rate. The inability of analysts to model estimated runs 'saved' on defense above or below the league average rate ("defensive runs") using traditional statistics led to the collection of proprietary trajectory and location data for every batted ball ("batted ball data"). Assuming such data is unbiased, reasonable estimates of defensive runs by fielders requires only a straightforward comparison of fielder success rates relative to the league in fielding batted balls in the same trajectory/location 'buckets'.

This presentation will describe the first method to estimate defensive runs (pitching and fielding) based solely on regression analysis applied to publicly available data. Seasonal team defensive statistics (strikeouts, shortstop assists, etc.) are centered relative to the league. Centered fielding plays are regressed upon similarly centered proxy variables for batted ball distributions. Residuals are treated as skill plays ("net plays"). A subsequent regression of team runs allowed relative to the league onto team net plays yield estimated runs per play. Runs per play multiplied by net plays yield runs per position, allocated among players based on plays relative to the team rate. Tests show resulting the defensive runs estimates correlate with published estimates based on batted ball data as well as the latter do with each other. See `www.oup.com/us/companion.websites/9780195397765/pdf/Appendix_A.pdf` for a complete disclosure.

A BRIEF STATISTICAL ANALYSIS OF LINE-UP IN THE MAJOR LEAGUE BASEBALL

Chulmin Kim†[1]

School of Mathematical Sciences, Rochester Institute of Technology[1]

† E-mail: *cxksma@rit.edu*

The line-up or the batting order in the baseball is the designated order in which the batters for a given team will bat. In the modern age of the Major league baseball, it has become more important to arrange the players into the right spots to produce offensive runs more efficiently. Even with the same nine batters for a given team, different batting order can make very different result in terms of producing runs. Are there any universal rules to decide a line-up? In general the managers probably place their better batters at the beginning spots of the line-up to give more chances for their teams and place the worse ones at the end of the line-up. Kim (2008) proposed a single offensive statistic K which measures a players overall batting ability. We study the relationship between the standardized Ks in the each line-up of a given teams and the teams winning percentage. And we characterize the line-up

empirically based on the data for the Major league baseball from the years 2001 to 2010 to use K and some other statistics. Those characteristics of the line-up are illustrated using a small simulated data set. The Wilcoxon-based rank statistics are used to analyze the line-up in the Major league baseball.

TESTING THE EFFICACY OF THE NBA'S AGE ELIGIBILITY RULE USING CENSORED REGRESSION

Ryan M. Rodenberg, JD/PhD[†1], Jun Woo Kim[1]

Department of Sport Management, Florida State University[1]

[†] E-mail: *rrodenberg@fsu.edu*

In 2005, the collective bargaining agreement (CBA) between the National Basketball Association (NBA) and the players union included, for the first time, a minimum age rule requiring all American-born players to be at least 19 years of age by December 31 of the draft year and be at least one year removed from high school in order to be draft eligible. Beck (2009) reported that the players union will seek to eliminate the age rule when the CBA expires in 2011, while Gregory (2008) quoted NBA Commissioner David Stern as saying that the league will move to increase the minimum age from 19 to 20. The purpose of this research is to clear the fog of any negotiation-related posturing and shed light on the interaction between age and career success. Specifically, we analyze the role of precocity on labor market outcomes of elite-level NBA players and, indirectly, test the on-court efficacy of the NBA's age rule. The resulting data set includes all first round NBA draft picks from 1989 to 2000. We show that precocity, as measured by the age of entry into the elite-level professional basketball labor pool, often leads to better career outcomes. There is no systematic evidence of any success among "late bloomers." Further, there is evidence that NBA team executives have, as a whole, accurately selected talented precocious players via the annual draft. Our findings cast doubt on the on-court efficacy of the NBA's contentious age eligibility rule.

A NEW PERFORMANCE MEASURE FOR LIMITED-OVERS CRICKET MATCHES

Aneesh Kulkarni[†1]

Microsoft Corporation[1]

[†] E-mail: *aneesh.kulkarni@post.harvard.edu*

The required run-rate has long been the most common measure for the strength of a team's position in the second innings. It is, unfortunately, a flawed measure as it fails to consider the batting team's wickets in hand. A fair method of combining the two resources at a batting team's disposal, overs remaining U and wickets in hand W, was proposed by Duckworth and Lewis in 1998, who modeled run scoring as a multivariate exponential function of U and W. We use the Duckworth-Lewis model to create a more accurate measure of the strength of a team's position during the second innings. We consider ratio of batting resources to the fraction of required runs remaining, and the overs remaining U as inputs to a logistic regression model to compute winnability values. Unlike required run-rate, winnability values can be directly compared across different stadiums and different teams. We show that on a large dataset of international Twenty20 matches, winnability is more accurate at predicting the outcome of games in progress than any heuristic based solely on required run-rate.

We also use winnability to derive a measure for the likelihood of a particular match outcome, given the strongest position achieved by the losing team. We show that during the 2009 IPL tournament, teams were losing matches from winning positions much more often than they would by chance.

"HEY REF, YOU OWE US ONE"

Mike Lopez†[1], Kevin Snyder[2]

Department of Biostatistics, Brown University[1]*, Sports Management Department, University of Massachusetts, Amherst*[2]

† E-mail: *Michael_Lopez@brown.edu*

In Game 5 of the 2011 Stanley Cup finals, the Boston Bruins were awarded the first four power plays of the contest, with the next and final three belonging to their opponents, the Vancouver Canucks. At the end of the game, the relatively even power play times suggest no evidence of official bias, but without any subconscious effort on behalf of the officials, how likely is this evening-up of penalty calls?

This project will focus on the number of penalty calls yielding power play minutes (thus excluding matching penalties) per team per period, which can roughly be approximated by a Poisson process with mean μ_{jk}, given period j (1,2,3) and team k (home, away). First, we'll estimate these means and standard errors using data from nhl.com boxscores, and, to consider the question of whether power play time tends to even out by the end of the game, calculate conditional means. In other words, given a higher, even, or lower number of penalties called on it in a previous period, can a team expect its mean number of penalties in the upcoming period to change? We'll answer this question, both based on the probability of

having a higher number of penalties called and in terms of the expected number of penalties called.

SUBOPTIMAL STRATEGIC DECISION-MAKING: EVIDENCE FROM THE NATIONAL FOOTBALL LEAGUE

Sean Macdonald†[1]

FTI Consulting[1]

† E-mail: *SMac2347@comcast.net*

This paper builds off of the work of Kovash and Levitt (2009) and evaluates managerial strategic decision-making, in the specific context of play calling by NFL offensive coaches. Minimax theory suggests that if coaches are behaving optimally, the different strategies that are chosen (e.g. run vs. pass) will have equivalent payoffs and the play-by-play selection of strategies will be independent. While Kovash and Levitt use five years of NFL regular-season data, I extend and update the analysis to a more recent eight year period and consider a broader range of play calling strategies. I estimate a model that predicts the value of having possession of the ball in specific circumstances, and then from this value function, calculate each play's value in terms of expected points. Next, I compare the relative value of different play calling options, considering the type of play, the play's direction, and the quarterback's alignment, to see if any strategy regularly outperforms another. I also look at serial correlation in play type selection from one snap to the next to discern if any exploitable pattern exists. I find that while there is little value difference in a play's direction, passes outperform runs; deep passes are superior to short passes; and having the quarterback in the shotgun is more valuable than having him under center. In addition, I find significant negative serial correlation in play calling. These findings illustrate that NFL coaches do not adhere to minimax theory, and thus in this respect are not optimal decision-makers.

A STATISTICAL LEARNING APPROACH TO EXTRACT THE BEHAVIORAL TRENDS IN SOCCER CLUB TEAMS RANKING POSITION

Abbas Saghaei†[1], Elaheh Nabati[1]

Industrial Engineering Department, Azad University[1]

† E-mail: *e.nabati@srbiau.ac.ir*

In this article changes in the ranking position of "Italian serie A" soccer teams are analyzed by applying a statistical learning frame work. Although teams' ranking analysis is a popular and important field, usually teams' rankings are compared with the previous results gained in one or a few time points. We consider team's alteration of the ranking position during a season as a curve; considering the data as functional objects (curves) provides additional and more precise information than single data points e.g. the speed and direction of team's performance change is easily detectable. The curves are gathered per time and per season. A curve clustering approach is implemented. The clustering method that is applied uses a novel algorithm that has recently been introduced and has not been studied in the field of sports so far. Similar trends are identified and grouped in the homogenous clusters. Afterwards association Rules method is applied to find which trends are relational, what the most frequent patterns are and which trend the next probable state of each team may be. Our proposed frame work provides practical results that are new and can be analyzed by sport specialists to help the sport clubs set their future policies.

EVOLUTIONARY MADNESS! APPLYING AN EVOLUTIONARY ALGORITHM TO OPTIMIZE TEAM SELECTIONS FOR AN NCAA TOURNAMENT CONTEST

Greg Chen, MA[1], Douglas Noe, PhD[†1]

Dept of Statistics, Miami University of Ohio[1]

[†] E-mail: *noeda@muohio.edu*

As one of the nation's most popular sporting events, the NCAA tournament always attracts millions of people's attention. In a typical "bracket" contest, people participate by filling in tournament brackets with the goal of correctly predicting the most round- and often upset-weighted game winners. We examine a unique contest structure in which each entrant selects a collection of teams subject to a seed-based cost constraint with the simple goal of maximizing the unweighted total number of wins.

We focus on one objective among a variety of team-selection strategies. Given a prior set of beliefs about the outcome probabilities for each potential tournament match, we seek to determine the set of teams with the greatest probability of achieving a preselected threshold number of wins. Due to the complex nature of the sample space, this problem is not easy to solve analytically. Instead, we apply an evolutionary algorithm to heuristically search the sample space by mimicking the process of natural evolution.

We examine the success of our approach using data from each NCAA men's basketball tournament from 2003 through 2011 and validate our algorithm using actual tournament results.

DO RULES OF TENNIS FAVOR THE BETTER PLAYER?

Reza Noubary†[1], Bill Calhoun[1]

Department of Mathematics, Bloomsburg University[1]

† E-mail: *rnoubary@bloomu.edu*

This article is an attempt to answer the question posed in the title. Using a simple model and an example we show that any reasonable scoring system with a large number of points will produce the same results. Our models predict that the use of a tie-breaker changes the outcome of a set very rarely. Even replacing the games with no-ad games, or sets with VASSS sets or matches with 150 point will hardly alter the outcome of a match in most cases. The model also shows that it is unlikely for the weaker player to win a match purely by chance. Thus if a player wins the match, she is very likely the better player, at least for that match.

GENERALIZED ODDS RATIOS FOR EVALUATING NHL PLAYERS BASED ON PLUS/MINUS STATISTICS

Douglas M. Okamoto, Ph.D.†[1]

Data to Information to Knowledge[1]

† E-mail: *d2i2k@aol.com*

National Hockey League (NHL) players are credited with a plus (+1) whenever their team scores an even strength or shorthanded goal, and debited with a minus (-1) whenever the opposing team scores an even strength goal while they are on the ice during a game. Although NHL players are not credited when their team scores on the power play, they are debited if their opponent scores a shorthanded goal while killing a penalty. After every game, a players pluses and minuses are tallied to get his plus/minus (+/-) statistic as reported in the box score of *Yahoo!Sports*, along with goals and assists. Since cumulative +/- statistics over an entire 82-game season measure offensive and defensive contributions, winning NHL teams have two-way players with highly positive +/- statistics. The odds ratio estimates the relative probability that a player has a positive +/- in a win, versus a negative +/- in a loss. If a player is on the ice for an even number of goals for and against his team (or none at all), his "zero" +/- is excluded from odds ratio estimation, as are shootout wins or losses when two teams remain tied after overtime. The purpose of this presentation is to estimate generalized odds ratios for the San Jose Sharks based on their plus/minus statistics

including zeroes, during the 2010-2011 regular season, as well as the post-season when they lost the Western Conference Finals of the 2011 NHL Stanley Cup Playoffs to the Vancouver Canucks.

LOGISTIC REGRESSION MODELS TO INVESTIGATE THE SUBTLE DIFFERENCES AMONG FIELD GOAL KICKERS IN THE NFL

Jason A. Osborne†[1]

Department of Statistics, North Carolina State University[1]

† E-mail: *osborne@stat.ncsu.edu*

Variability of performance among field goal kickers in the National Football League is surprisingly small. In several studies (e.g. Morrison and Kulwani, 1993), the null hypothesis of equal accuracy among all kickers, after accounting for variability in the distance of the attempt, was not rejected. In this paper we assemble a dataset of more than 5000 field goal attempts over the seasons from 2005-2010 and revisit the question. (We also provide guidance on how to use PERL to retrieve the data from espn.com.) We are able to identify modest differences, mostly due to a few bad kickers. These differences are modeled by comparing kicker-specific slopes in logistic regression models. We also investigate the association between accuracy and the position at which the kicker was selected in the NFL draft, and whether or not the player was drafted at all.

TEAM BASED SHOT AND SCORE PREDICTION IN NBA

Jung-Ho Park[1], Anthony Bedford†[1]

School of Mathematical & Geospatial Sciences, RMIT University, Melbourne, VIC, Australia[1]

† E-mail: *anthony.bedford@rmit.edu.au*

Score prediction in any sport is the epitome of the sports statisticians' skills. The prediction of score both pre and in game remains a task to these soothsayers. Notably, bookmakers are very good in their ability to predict the line, and this information provides a useful starter for our work. In our attempt to predict NBA final score, we utilise a three-stage process to ultimately determine the final score for each team in each match during the

season. We test the model's efficiency in-the-run and look at team-based variation in results. We model, using a series of regression equations, the relationship between shot attempts (including free-throws), game pace, and score, and examine the correlation between attempts and score in-game. This reveals to us clear differences in teams' approaches to scoring, and the relationships between team-based game pace and scoring patterns. In addition, we investigate the score distribution by analysing the relationship between the total shots attempted and shot percentage variation within live basketball matches. These results inform us of potential opportunities for wagering. We also explore the idea of the home team setting the tempo, that is, does the home team moderate the tempo.

THE IMPACT OF BALL POSSESSION IN PROFESSIONAL SOCCER ON GOAL SCORING

Eric Parziale[1], Philip Yates[†1]

Saint Michael's College[1]

[†] E-mail: *pyates@smcvt.edu*

In this study, play-by-play data from the English Premier League and Brazil's first and second division were obtained from StatDNA. Via logistic regression, various strategies (short versus long passes along the ground, short number of touches on the ball, passes in different thirds of the field, etc.) were investigated to see how likely a player will successfully complete a pass and ensure his team possesses the ball. Multilevel logistic regression was used to identify which players and teams are most successful in terms of pass completion in various in-game situations. The play-by-play data were summarized into game-by-game data, and Poisson regression was used to summarize the relationship between ball possession and margin of victory.

GAUSSIAN PROCESSES FOR PREDICTING THE OUTCOME OF TENNIS MATCHES

Javier M. Cespedes[1], Fernando Perez-Cruz[†1]

Signal Theory and Communications Department, University Carlos III, Madrid, Spain[1]

[†] E-mail: *fernando@tsc.uc3m.es*

We have applied Gaussian Processes for Machine Learning to predict the outcome of tennis matches using 14 different features (six describing each player and two about their previous

matches). Gaussian Process is a Bayesian tool to learn a regression function or a classifier. It assumes that the observe output comes from an underlying Gaussian Process and an observation model, which models if the problem is a classification (discrete output) versus regression (continuous output). Gaussian processes can capture nonlinear interaction in the data using a nonlinear covariance function, whose hyper-parameters can be learnt by maximum likelihood, and they provide accurate posterior probability estimates.

For predicting the tennis matches we have used 14 features. First we put the values for the player with highest ranking and then for lowest ranking player in the game and we measure:
Points won serve (percentage)
Points won rest (percentage)
Total number on games in the last 30 days.
Total number of games won in the last 15 games.
Total number of games won in the last 15 games in the surface the game is being played.
First service for the last 15 games (percentage)

Finally we add the number of games that player with highest ranking has won to the other player in the same surface and vice versa. The resulting model is able to predict more accurately the outcome of the game for 2011 clay season, when compared to the posted betting odds by different online betting parlors.

ESTIMATING FIELDING ABILITY IN BASEBALL PLAYERS OVER TIME

James Piette[†1], Shane T. Jensen[2]

Krossover Intelligence Inc.[1]*, Department of Statistics, Wharton School of Business*[2]

[†] E-mail: *jamesmpiette@gmail.com*

It is commonplace around baseball to involve statistical analysis in the evaluation of a player's ability to field. While well-researched, the question behind how best to model fielding is heavily debated. The official MLB method for tracking fielding is riddled with biases and censoring problems, while more recent approaches to fielding evaluation, such as Ultimate Zone Rating [Lichtman 2003], lose accuracy by not treating the field as a single continuous surface. SAFE, Spatial Aggregate Fielding Evaluation, aims to solve these problems. Jensen et al [2009] took a rigorous statistical approach to this problem by implementing a hierarchical Bayesian structure in a spatial model setting. The performance of individual fielders can be more accurately gauged because of the additional information provided via sharing across fielders. We have extended this model to three new specifications by building in time series aspects: the constant ability model, the moving average age model and the autoregressive

age model. By using these new models, we have produced a more accurate estimation of a player's seasonal fielding performance and added insight into the aging process of a baseball player's underlying ability to field.

IN-GAME WIN PROBABILITY FOR BASKETBALL

Jason Rosenfeld[†1], Carl Morris[1]

Dept of Statistics, Harvard University[1]

[†] E-mail: *Jasonwrosenfeld@gmail.com*

We review and develop ways to calculate in-game win probabilities (IGWP) during a basketball game. One way involves approximation by Normal distributions, justified by the central limit theorem until late in games. However, modeling late in games becomes challenging because constant variance over fixed time intervals fails in close games when trailing teams adopt tactics that increase the number of possessions (and so, variability) per unit time, while their opponents do the opposite. Other IGWP formulas can be based on logistic modeling of within-game data. Some analyses have ignored measures of team strengths (perhaps using home court advantage as a substitute), while point-spread information and team records provide valuable improvements. We assess IGWP formulas in light of data for several seasons of NBA games.

Valid IGWP measures can aid teams in several ways. They can help coaches gauge when it is safe to rest their best players, and also support in-game strategic decision-making. For post-game assessments, they help identify a game's key turning points. Plus/minus (+/-) measures ordinarily are based on point differentials while a player is on the floor. IGWP formulas can suggest how plus/minus weights should be modified during garbage time and during crunch time. Plus/minus measures also can be based on IGWP differentials.

USING AN OPTIMIZED EXPONENTIAL MODEL FOR WAGERING ON LIMITED-OVERS CRICKET MATCHES

Jonathan Sargent[†1], Anthony Bedford[1]

School of Mathematical & Geospatial Sciences, RMIT University, Melbourne, VIC, Australia[1]

[†] E-mail: *jonathan.sargent@student.rmit.edu.au*

This paper introduces an exponential player-based model that recalculates in-game odds during limited-overs cricket matches for the estimated period leading up to a batters dismissal. This period is derived from a Bayesian estimate of the number of runs the batter is expected to achieve – weighted by his and the bowlers player ratings – with his teams odds lengthening as he approaches actual and expected run convergence. Punters are, hence, offered an arbitrage occasion because the market price reaction occurs only after the batter is dismissed. Using three years of match data, a teams runs (r), wickets lost (w), opponent strength difference and home-ground advantage were applied in a logistic function to arrive at an in-game win probability, which was converted to a match price at time t with current resources (r_t, w_t). A subsequent win probability was pre-determined for the loss of the next wicket, or resources $(r_t, w_t + 1)$, and an exponential price curve fitted between the two. The growth rate of the curve was optimized to maximise the return on investment, calculated as the positive difference between the models odds at time t and pre-collected ball-by-ball market odds. Using a recent limited-overs cricket match as a case study, this paper will demonstrate that the models ability to adjust odds in anticipation of a wicket, rather than in reaction to one, improves return on investment when betting during a match. Furthermore, the model would simply translate to other bat-and-ball sports, such as baseball.

RANKING PROFESSIONAL GOLF PLAYERS – WHO IS NUMBER 1?

Kabir C. Sen[†1], Charles F. Hawkins[1]

Lamar University[1]

[†] E-mail: *Kabir.Sen@Lamar.edu*

In our competitive, market oriented world, it is almost impossible not to encounter the question "Who is No. 1?" on a daily basis. In the golfing world, several polls or ranking schemes have been developed to rank professional golfers. Among these are PGA TOUR official earnings, Official World Golf Ranking, PGA Tour rankings, and FedEx Cup rankings. One of the criticisms of the current ranking schemes is that they do not take into consideration a player's failure to participate in some officially sanctioned tournaments. Also, the time span for looking at performances can vary across the different types of evaluations. This paper will compare different rankings of PGA Tour golfers and take into account the participation and time span factors. The paper will suggest some changes that a new system of computing World Golf rankings might incorporate in order to accurately reflect the latest form of a golfer.

RISK ANALYSIS OF PLAYER DECISIONS IN SOCCER

Sawyer Tabony†[1], Anna Teytelman[2], Mayank Varia[3]

Boston College[1], MIT[2], Lincoln Labs[3]

† E-mail: *sawyertabony@gmail.com*

In this work, we analyze risk in decision making by players during a soccer match. Often, a player must choose between a risk-averse strategy (such as a defender pressuring the ball) and a risk-prone strategy (slide-tackling to win the ball back). We use the data from StatDNA to determine the riskiness of various situations on the field by finding the probabilistic consequences of an instance of the actions. Also, we characterize players by their risk-averse or risk-prone tendencies.

We generally characterize the risk tendencies of attackers and defenders, determine specifics about the most common errors players make in managing risk, and examine which types of players work effectively together. Our analysis could help coaches evaluate or improve the decision-making capabilities of their players, or assist their tactics by better controlling risk on the field.

PREDICTIVE VALUE OF NFL COMBINE ON FIELD PERFORMANCE

Masaru Termoto†[1], Chad L. Cross[2]

Kinesiology Department, Wesley College[1], Crossroads Wellness, LLC[2]

† E-mail: *Masaru.Teramoto@wesley.edu*

The National Football League (NFL) Scouting Combine invites college football players every year to assess various athletic abilities. Although the combine provides NFL coaches and scouts with an opportunity to evaluate college players for the upcoming NFL draft, its value on predicting future success of players has been questioned. The aim of this study is to examine whether the NFL Combine performance can predict future performance of running backs (RBs) and wide receivers (WRs). We analyzed the combine data of RBs and WRs in the past 10 years and their field performance in the NFL using a weighted least squares regression analysis. The results of the data analysis indicate that, after adjusting for the number of games played, the 10-yard dash performance is significantly associated with the career rushing yards per attempt in RBs. In addition, there is a tendency that RBs with

faster times on 3 cone drill have better career rushing yards per attempt. In WRs, height and vertical jump are the significant predictors of the career receiving yards per reception. It appears that the NFL Combine has some value for predicting future performance of RBs and WRs.

HOW TO GET FROM PITCH TO PLINTH: MODELLING STATUE SUBJECT CHOICE IN US BASEBALL AND ENGLISH SOCCER

Ffion Thomas[†1], Chris Stride[1], Josh Pahigian[2]

University of Sheffield, UK[1]*, University of New England, USA*[2]

[†] E-mail: *ega07fet@sheffield.ac.uk*

From an almost standing start in the late 1980s, the number of statues of US baseball players and UK soccer players has risen inexorably. By April 2011, almost 50 soccer statues and over 130 such baseball statues were in situ. Yet even amongst the best players, few will ever be chosen as subjects.

We investigated and compared the defining characteristics of statue subjects in these two national sports. The MLB Hall of Fame and the English Football League 100 Legends list were used as samples of the best performers from each sport; approximately half of the baseball players and a third of soccer players from these lists have been depicted in bronze. Logistic regression models were built to predict choice as a statue subject, with hypothesised predictors including loyalty (% career matches played for primary team), locality (was player born locally to primary team), career longevity, performance statistics of player and primary club, national recognition (international/all-star team selection), sympathy (early death) and the effect of nostalgia/memory in selecting statue subjects (i.e. the era a player performed in).

Results for both sports indicate the importance of club loyalty, and the dominant effect of era. Players from the 1950's-1970's were most likely to be depicted, suggesting that the role of a statue as a nostalgia/heritage marketing object impacts upon subject choice. Players deserving depiction if the "chance-of-birth" effect of era is removed were identified, and predicted probabilities of depiction were calculated for players with statues who were not part of the legends' samples.