PROGRAM

September 26, 2009

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

2009 New England Symposium on Statistics in Sports

September 26, 2009

Symposium Co-Organizers:

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

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

Sponsors:

- Boston Chapter of the American Statistical Association
- Section on Statistics in Sports of the American Statistical Association
- ESPN
- Harvard University Department of Statistics

Acknowledgments: We wish to thank everyone who helped to make this event possible. Special thanks go in particular to Tom Lane for creating and maintaining the NESSIS web site and sending out BCASA announcements, Jason Rosenfeld for helping to find speakers and panelists and for publicizing the event to other students, Stephen Rutkowski for enabling ESPN to contribute funding to this event, Ben Alamar for creating a special NESSIS issue of the Journal of Quantitative Analysis in Sports, and Mike Zarren for arranging a post-conference get-together at *The Fours* and contacting potential speakers. We would also like to thank James O'Malley, Phil Everson, Betsey Cogswell, Ellen Weene, Dale Rinkel, Carl Morris, Iram Farooq, and Huichao Chen for their parts in helping with the conference.

2009 New England Symposium on Statistics in Sports

September 26, 2009

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

9:00am – 9:15am:	Welcome
9:15am – 9:45am:	Benjamin Alamar, Menlo College *"Evaluating Risk in NFL Play Calling"*
9:45am – 10:15am:	Jeffrey L. Cornett, Valencia Community College *"Visualization of Crew Race Performance: Drives and Duels"*
10:15am – 10:45am:	Wayne Winston, Indiana University *"Player and Lineup Analysis in the NBA"*
10:45am – 11:00am:	Break (coffee and tea)
11:00am – 11:30am:	Radu Tunaru, City University London *"Valuations of Soccer Players from Statistical Performance Data"*
11:30am – 12:00pm:	Tom Tippett, Boston Red Sox *"Using Lineup-Dependent Expected Runs Analysis to Evaluate Baseball Tactics"*
12:00pm – 1:00pm:	Lunch break
1:00pm – 1:20pm:	Blakeley McShane, Wharton School *"Evaluating New Pitching Metrics using a Point-Mass Mixture Random Effects Model"*
1:20pm – 1:40pm:	Maggie Wigness, Pacific University *"Using New Iterative Methods and Fine Grain Data to Rank College Football Teams"*
1:40pm – 2:00pm:	Daniel Porter, Columbia University *"Do Hitters and Pitchers Vary in their Sensitivities to Changes in the Quality of the Opposition?"*
2:00pm – 2:15pm:	Break

2:15pm – 2:45pm: Wayne DeSarbo, Pennsylvania State University
"Examining Heterogeneous Expressions of Sports Fan Avidity: a Spatial Multidimensional Scaling Approach"

2:45pm – 3:15pm: Gil Fellingham, Brigham Young University
"Skill Importance in Women's Volleyball"

3:15pm – 4:45pm: Poster Session, with snacks and beverages

4:45pm – 6:00pm: Panel Discussion – *"Data and decisions in basketball: A peek into the minds of NBA statisticians."*

Moderator: Greg Dickerson – Boston Celtics sideline reporter for CSN New England

Panelist: Mike Zarren – Boston Celtics
Panelist: Ken Catanella – National Basketball Association
Panelist: Aaron Barzilai – Memphis Grizzlies

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

Oral Presentation Abstracts

EVALUATING RISK IN NFL PLAY CALLING

Benjamin Alamar[†]

Menlo College

[†] E-mail: *quantsports@gmail.com*

Coaches in the NFL make approximately 1000 offensive play calls during the regular season. These calls are the result of countless hours of preparation and analysis and the coach's own personal experience and each coach has their own measures of success and biases regarding types of play calls. What has not been utilized previously, is a systematic analytical approach to measuring a play's outcome in relation to the drive, and an evaluation of whether coaches' are irrationally biased in their playcalling. Using play by play data from the 2005 through 2008 NFL regular season, an evaluation system is built around the concept of expected points. Expected points have been used in baseball for over 40 years and have been applied occasionally in football (Romer 2003; Carroll et al 1988), for all downs, with no increased risk. These results confirm previous analysis (Alamar 2008) that teams underutilize the pass. Additionally, the expected points framework was applied to every team in the league for each season, to determine the best offenses and defenses (overall and broken down by pass/run) during the time span of 2005-2008.

VISUALIZATION OF CREW RACE PERFORMANCE: DRIVES AND DUELS

Jeffrey L. Cornett[†]

Managing Director of Institutional Research, Valencia Community College

[†] E-mail: *jcornett1@valenciacc.edu*

The visualization and modeling of crew race strategy is limited by the sparse information tracked and published on crew team performance. Video records of rowing world championships provide a rich source of data for those capable and patient enough to mine this level of detail. This research is based upon detailed frame-by-frame video analysis of five

world championship rowing finals. The patterns of drives that crews make on each other is graphically illustrated, with particular emphasis on the position duels that proved decisive in determining the order of finish. Conditional probabilities of winning are computed as a function of location on the course, proximity to opponents, and whether a crew is driving on its opposition. This data enables the computation of what constitutes a probabilistically decisive lead during a race, thus providing real time racing guidance as to when crews should consider altering their race goals and strategy.

This presentation introduces new concepts into the vocabulary of race strategy and performance assessment including position duels, drive model of a duel, challenge vs. decisive drives, and probabilistically decisive leads. These concepts are theoretically applicable to any athletic race competition, but are especially applicable and easy to graphically illustrate for the sport of rowing. In rowing, the pacing of drives that crews make on each other is relatively slow, thus allowing race participants to observe and potentially react to unexpected race situations.

EXAMINING HETEROGENEOUS EXPRESSIONS OF SPORTS FAN AVIDITY: A SPATIAL MULTIDIMENSIONAL SCALING APPROACH

Wayne S. DeSarbo[†1], Robert Madrigal[2], Robert Fox[3]

[1] *Pennsylvania State University,* [2] *University of Oregon,* [3] *TNS Sport and Sponsorship*

[†] E-mail: *wsd6@psu.edu*

Fan avidity refers to the level of interest, involvement, intensity, passion, and loyalty a fan gives to a particular sport, league, and/or team. This concept involves a fan.s allocation of time and/or expenditures towards a given sport. As such, fan avidity has been measured by sports marketing researchers in terms of a variety of different survey items related to interest, involvement, viewership, merchandise/ ticket purchases, etc. As numerous researchers have found, fan avidity can be expressed by fans in a number of different ways depending upon the specific fan, sport, and team involved. The objective of this research is to uncover the dimensional aspects of such manifestations of fan avidity and explore the nature of heterogeneity in such expressions. We focus on the student fans of a well known Big Ten university football program and develop some 35 different expressions of fan avidity related to how student fans follow their team. We develop a spatial choice multidimensional scaling model and uncover four latent dimensions of fan avidity expression. We illustrate how simple fan avidity measurements using interest and involvement scales are insufficient to fully capture such heterogeneous expressions, and conclude that academics and practitioners need to enhance current fan avidity measurement procedures.

SKILL IMPORTANCE IN WOMEN'S VOLLEYBALL

Michelle Allan, Gilbert Fellingham[†], Lindsay Florence

Brigham Young University

[†] E-mail: *gwf@byu.edu*

The purpose of this paper is to demonstrate two methods to quantify skill importance for teams in general, and women's volleyball in particular. A division I women's volleyball team rated each skill (serve, pass, set, etc.) and recorded rally outcomes during all home games in a competitive season. The skills were only rated when the ball was on the home team's side of the net. Events followed one of these three patterns: serve-outcome, pass-set-attack-outcome, or block-dig-set-attack-outcome. These sequences of events were assumed to be first-order Markov chains, meaning the quality of the performance of the current skill only depended on the quality of the performance of the previous skill. We analyze the volleyball data using two different techniques: one uses a Markovian transition matrix, while the other is an implementation of Bayesian logistic regression. To estimate the Markovian transition matrix, we assumed a multinomial likelihood with a Dirichlet prior on the transition probabilities. The logistic regression model also uses a Bayesian approach to determine how the performance of the skill affects the probability of a successful outcome. The posterior distributions of parameters associated with skill performance are used to calculate importance scores. Importance scores produced by the two methods are reasonably consistent across skills. The importance scores indicate, among other things, that the team would have been well rewarded by improving transition offense. Importance scores can be used to assist coaches in allocating practice time, developing new strategies, and optimizing team performance relative to player selection.

USING LINEUP-DEPENDENT EXPECTED RUNS ANALYSIS TO EVALUATE BASEBALL TACTICS

Tom Tippett[†]

Boston Red Sox

[†] E-mail: *ttippett@redsox.com*

A common method for evaluating baseball tactics is to compare the increase in the expected number of runs that will score if the attempt succeeds with the loss of run-scoring potential

if it fails. Using these two values, one can then say that the chances for success must exceed a certain percentage to make the attempt worthwhile.

In today's high-offense era, this approach often leads to the conclusion that certain tactics don't make sense even if success is guaranteed because the cost of the out given up is always greater than the value of the base(s) gained. Yet these tactics are still commonly employed in today's brand of baseball.

But the conclusion that many of these tactics are outmoded is based on expected run values that are derived from league averages. As a result, those values assume that a sequence of average hitters is due up after the tactic has been employed.

For this presentation, we will describe a method for computing expected runs values for specific sequences of hitters. Using that method, we will evaluate tactics such as the sacrifice bunt and assess whether the quality of the opposing pitcher and the nature of the ballpark should play a major role in making these decisions.

VALUATIONS OF SOCCER PLAYERS FROM STATISTICAL PERFORMANCE DATA

Radu Tunaru[†1], Howard Viney[2]

[1] *City University London,* [2] *Open University*

[†] E-mail: *R.Tunaru@city.ac.uk*

Based upon contingent claims methodology and standard techniques in statistical modeling and stochastic calculus, we develop a framework for determining the financial value of professional soccer players to their existing and potential new clubs. The model recognizes that a player's value is a product of a variety of factors, some of them more obvious (i.e. on-field performance, injuries, disciplinary record), and some of them less obvious (i.e. image rights or personal background). We provide numerical examples based upon historical statistical performance indicators that suggest the value of a soccer player is not the same for all potential clubs present in a market. In other words this is a special case where the law of one price for one asset does not function. Our modeling employs the vast database of soccer players' performance maintained by OPTA Sportsdata Ltd, a subsidiary of SkySports the UK broadcaster; the same database has been used by major clubs in the English Premiership such as Arsenal and Chelsea. From a statistical point of view, our model can be applied to identify the relative value of players with similar characteristics but different market valuations, to explore patterns of performance for individual star players and teams over a run of games, and to explore correlations or interactions between pairs of players or small group of players in the team. Moreover, it offers a tool to value players from a financial point of

view using their past performance; hence this model can be also used to inform contractual negotiations.

PLAYER AND LINEUP ANALYSIS IN THE NBA

Wayne Winston[†]

Kelley School of Business, Indiana University

[†] E-mail: *winston@indiana.edu*

We will discuss our nine years of working for the Dallas Mavericks. We will begin by reviewing methods for rating players and then discuss how to rate lineups and analyze matchups during the NBA playoffs. We will discuss several surprising findings. For example, Lamar Odom was a more effective player during the 2009 playoffs than Kobe Bryant.

Student Oral Presentation Abstracts

EVALUATING BASEBALL METRICS USING A POINT-MASS MIXTURE RANDOM EFFECTS MODEL

Blakeley B. McShane[†], James Piette, Shane T. Jensen

Department of Statistics, Wharton School, University of Pennsylvania

[†] E-mail: *mcshaneb@wharton.upenn.edu*

The vast proliferation in data and analyses made possible by technological advances over the past decade has provoked a new problem: with so many variables being tracked on individuals throughout time, which ones demonstrate signal and which are noise? We address this pressing issue by adapting existing variable selection approaches to create a Bayesian random effects model with a point mass mixture and use Major League Baseball data as a case study. With the rise of Moneyball and sabermetrics, traditional measures of performance (e.g., ERA, wins for pitchers; AVG, RBIs for hitters) have been supplanted by new ones (e.g., WHIP, strikeout to walk ratio for pitchers; OBP, BABIP for hitters). For pitching, our findings confirm baseball analysts' speculation: the metrics that show the most signal

are those that gauge outcomes which are independent of a team's defense (e.g., strikeouts per nine innings and groundball percentage). For hitting, almost all metrics show signal but that is because they are highly correlated with traditional measures of power, speed, ability to make contact, and plate discipline. We recommend that sabermetricians concentrate their analyses on new pitching metrics and in particular fielding independent metrics, develop new ones orthogonal to extant ones, and shy away from those outcomes which depend on fielding.

DO HITTERS AND PITCHERS VARY IN THEIR SENSITIVITIES TO CHANGES IN THE QUALITY OF THE OPPOSITION?

Daniel Porter[†]

Columbia University

[†] E-mail: *danielwporter@gmail.com*

This study quantifies the extent to which the actual performance of a certain batter versus a certain pitcher should be weighted against the overall performance of each player. A hierarchical model is used in order to properly account for varying sample sizes in terms of measuring tendencies in batter versus pitcher matchups. It builds upon the approach introduced by Stern and Sugano (2007), but incorporates additional information about each player into forming the prior distribution of the hyperparameters. Such an approach provides a posterior outcome with increased predictive value, the implications of which can have practical value for decision makers within the game. By allowing increased inference into the sensitivity of batter and pitcher performance to the quality and handedness of the opposition, teams can optimize playing time based upon the comparative advantage of its various players, and properly evaluate the relative values of different players in high leverage (i.e. postseason) situations when considering a particular acquisition.

USING NEW ITERATIVE METHODS AND FINE GRAIN DATA TO RANK COLLEGE FOOTBALL TEAMS

Maggie Wigness[†], Michael Rowell, Chadd Williams

Pacific University

[†] E-mail: *wign5556@pacificu.edu*

At the end of each NCAA Division I-A football season, nearly half of the teams will compete in a bowl game looking for one final victory. The Bowl Championship Series (BCS) ranking system plays a significant role during this part of the season, as it determines which two teams will compete in the BCS National Championship Game, helps break conference ties, and influences the selections of smaller bowl games. Due to mispredicted bowl games each year, confidence in the BCS rankings is low. This paper makes three main contributions to address the validity of this concern. First, using the results of past bowl games, we analyze the accuracy of the BCS computer rankings, assuming that the team receiving a higher ranking should win the bowl game. Second, using a basic set of game statistics that include wins/losses, final scores, location of games, and strength of schedule, we devise new iterative ranking methods using mathematical modeling and graph theory. Third, in an effort to deviate from the typical approach of ranking college football teams, we analyze a variety of play-by-play statistics as an indicator of team strength. Using this analysis we devise a new ranking method using this fine grain data. Finally, each new method is analyzed in the same fashion as the BCS computer rankings, and the accuracy is compared to that of the BCS method.

Poster Presentation Abstracts

QUANTIFYING THE EFFECT OF PERFORMANCE-ENHANCING DRUG USE ON FASTBALL VELOCITY IN MAJOR LEAGUE BASEBALL

Vittorio Addona[†], Jeremy Roth

Macalester College

[†] E-mail: *addona@macalester.edu*

Since 2005, Major League Baseball (MLB) has suspended 258 players under its Drug Prevention and Treatment Program. Moreover, the Mitchell Report yielded the names of 89 alleged users of performance-enhancing drugs (PEDs). This documentation enables quantification of the impact of PEDs on player performance. Literature with this goal is limited, and has focused primarily on batters. Some authors have examined Roger Clemens, but there has been no previous work assessing the influence of PEDs on pitchers more generally. We gathered average fastball velocity from Fangraphs.com for all MLB pitchers who threw at least 10 innings in a month between 2002 and 2008 (11,860 player months). Pitchers

were deemed to be PED users if they were named as such in the Mitchell Report or suspended by MLB for a positive PED test. Human growth hormone (HGH) usage was tracked separately. We modeled fastball velocity by PED and HGH usage, age, a Starter/Reliever indicator, and several control variables. Using PEDs increased average fastball velocity by 1.074 MPH ($p < 2 \times 10^{-16}$) overall. When PED impact was allowed to vary by pitcher type (Starter/Reliever) and age, its benefits were most substantial later in a player's career. For example, at age 35, the effect of PEDs was 1.437 MPH ($p < 2 \times 10^{-16}$) for relievers and 0.988 MPH ($p < 2 \times 10^{-16}$) for starters. HGH use was significantly negatively correlated with fastball velocity. This suggests disproportional HGH use by injured players hoping to hasten their recoveries, and is consistent with frequent explanations provided in the Mitchell Report.

DECIDING A CRICKET MATCH AFTER INTERRUPTIONS

Chaithanya Bandi[†], Michael Frankovich

Operations Research Center, Massachusetts Institute of Technology

[†] E-mail: *cbandi@mit.edu*

One-day Cricket is a popular game internationally, being played most notably by the major countries of the Commonwealth, and in particular India, where the game draws an extremely large, devoted fan base and significant financial interest. When rain or other interruptions occur, the duration of play of a game is often shortened in order to complete the game within the allotted time frame, typically eight hours. In such cases it is normally necessary to adjust the target score of the team which bats second in order to mitigate the effects of the interruption. This can favor one side or the other, depending on the circumstances. While many methods have been proposed, the Duckworth/Lewis method (D/L) has been officially used in One-Day International games since 2001 and seems to produce reasonable results. We propose an improved method which, while similar to D/L, takes additional considerations into account. In particular, we allow the value function of expected runs, given a particular state, to depend on the quality of a team, as well as momentum. Computational experiments using real world data indicate that the method produces comparable results to D/L, and many practical instances are shown where the method gives more meaningful results. Intuition is provided why the method is a strict improvement over D/L Method.

USING RANDOM FORESTS TO PREDICT BASEBALL PLAYERS' PROBABILITIES OF ELECTION TO THE HALL OF FAME

Michael Freiman[†]

Department of Statistics, Wharton School, University of Pennsylvania

[†] E-mail: *mfreiman@wharton.upenn.edu*

A popular topic of argument among baseball fans is the prospective Hall of Fame status of current and recently retired players. The problem of estimating a player's probability of enshrinement can be approached by machine learning methods, and in particular I consider the use of random forests for this purpose. A random forest may be considered a black-box method for predicting Hall of Fame status, but a number of parameters must be chosen before the forest can be grown. These parameters include fundamental aspects of the nuts and bolts of the construction of the trees that make up the forest, as well as choices among possible predictor variables. For example, one predictor that may be considered is a measure of the player's having seasons with many home runs hit, and there may be multiple competing ways of measuring this. Furthermore, certain methods of searching the parameter space are made less useful by the randomness underlying the forest's construction and the fact that, by sheer luck, two forests constructed with the same parameters may have differing qualities of fit. I examine how a stochastic method such as simulated annealing can be used to try to find forests that solve the problem of the "lucky forest."

BAYESBALL: A BAYESIAN MODEL FOR EVALUATING FIELDING IN MAJOR LEAGUE BASEBALL

Shane Jensen[†]

Department of Statistics, Wharton School, University of Pennsylvania

[†] E-mail: *stjensen@wharton.upenn.edu*

We present sophisticated methodology for the analysis of fielding performance in major league baseball. Our approach is based upon high-resolution data consisting of on-field location of batted balls. A key issue is the balance between the personal performance of an individual fielder and the shrinkage to the population performance of similar fielders. We combine spatial modeling with a hierarchical Bayesian structure in order to evaluate individual fielders while sharing information between fielders at each position. We present results

across four seasons of MLB data and compare our approach to other fielding evaluation procedures.

A MODIFICATION OF OPS: WIDELY USED TO MEASURE A BASEBALL BATTER'S PERFORMANCE

Chulmin Kim[†]

School of Mathematical Sciences, Rochester Institute of Technology

[†] E-mail: *cxksma@rit.edu*

Batting average (BA), home runs (HRs), and runs batted in (RBIs) and have been the most dominant statistics to measure a baseball batter's performance. Since each of those three contains a meaningful interpretation but also some drawback to explain a batter's ability at the same time, often time we see the three together in baseball broadcasting or sports sections of newspapers. Slugging percentage (SLG) and on-base percentage (OBP) have been used as alternatives of the traditional three statistics. SLG measures how often a batter hits and how valuable the hits are and OBP measures how often a batter reaches bases. Whereas SLG ignores reaching bases by hits by pitched ball or walks, OBP is limited to measure the quality of the hits. A combination of these two is called OPS, the sum of OBP and SLG, which has become more widely used. There have been studied by several slightly varied statistics of OPS. One example of those is called gross point average (GPA) introduced by Gleeman. GPA is obtained by the following: (the sum of SLG and 1.8 times OBP)/4. We introduced a variation of these statistics, K, which is a single number explaining not only a batter's hitting performance but also his non-hitting performance to generate runs for his team such as stolen bases, walks, and etc. This newly developed statistic is based on major league team statistics from the year 2000 to the year 2008. We propose a model to a baseball batter's salary by a measure of his contribution to generate runs for his team, K.

ROC CURVES TO MEASURE THE QUALITY OF THE DECISIONS IN CRICKET

Ananda Manage[†1], Kumudu Mallawarachchi[2]

[1] *Department of Mathematics and Statistics, Sam Houston State Univ;* [2] *Department of Mathematics and Statistics, Texas Tech University*

[†] E-mail: *abandulasiri@shsu.edu*

A Receiver Operating Characteristic (ROC) curve visually demonstrates the tradeoff between sensitivity and specificity as a function of varying a classification threshold. It is a common practice to use ROC curves to measure the accuracy of predictions by different methods. Although this method has been used primarily in medical and engineering fields, it could be used effectively in sports as well. More precisely, an ROC plots the sensitivity versus (1 – specificity), and the area under the curve gives a measure of the prediction. So, the ideal best prediction should have one square unit of area under the ROC, where it achieves both 100% sensitivity and 100% specificity (which, in reality, rarely happens). Consequently, when we compare two methods, the one with the greater area under its ROC is judged best. This paper shows the effectiveness of using the ROC curves in analyzing cricket data. In particular, the quality of the decisions made by umpires will be investigated. And also, the comparison of the accuracy of the Duckworth-Lewis method for revising target for matches shortened due to weather interruptions will be key interests in our investigation.

ANOMALIES IN TOURNAMENT DESIGN: THE MADNESS OF MARCH MADNESS

Robert Baumann, Victor Matheson[†], Cara Howe

College of the Holy Cross

[†] E-mail: *vmatheso@holycross.edu*

Tournament design is of crucial importance in competitive sports. The primary goal of effective tournament design is to provide incentives for the participants to maximize their performance both during the tournament and in the time period leading up to the tournament. In spectator sports, a secondary goal of tournament design is to also promote interesting match ups that generate fan interest. Seeded tournaments, in general, promote both goals. Teams or individuals with strong performances leading up to a tournament receive higher seeds which increase their chances of progressing further in the tournament. Furthermore, seeding ensures that the strongest teams or players are most likely to meet in the final rounds of the tournament during which time fan interest is at its peak. Under some distributions of team or player skill, however, a seeding system can introduce anomalies that could affect incentives.

Our analysis of the NCAA men's basketball tournament uncovers such an anomaly. The seeding system in this tournament gives teams with better success in the regular season more favorable first round match ups, but the tournament is not reseeded as the games progress. Therefore, while higher seeds progress to the 2nd round of the tournament at uniformly higher rates than lower seeds, this relationship breaks down in later rounds. We find that 10th and 11th seeds average more wins and typically progress farther in the tournament

than 8th and 9th seeds. This finding violates the intended incentive structure of seeded tournaments.

ODDS RATIO MODELS IN SPORTS

Carl Morris[†]

Department of Statistics, Harvard University

[†] E-mail: *morris@stat.harvard.edu*

Odds ratio and logistic models arise naturally in paired comparison settings. That makes them very useful in sports to explain and model and predict winning and other successes and failures. These models emerge from theoretical considerations in some sports examples. More generally, their applications concern baseball (e.g., Bill James' "Pythagorean" formula), basketball, and tennis, and how different estimates of model parameters can stem from different aspects of the data for fitting these models. Related issues include developing standard deviations for such models.

TAIL MODELING, TRACK AND FIELD RECORDS, AND USAIN BOLT

Reza D. Noubary[†1]

Department of Mathematics and Statistics, Bloomsburg University

[†] E-mail: *rnoubary@bloomu.edu*

Athletic performances and related issues have received considerable attention by physiologists, physical educators, and public. One aspect is concerned with the improvement of the records over time in order to address the question of forecasting the future records. This article presents a procedure for calculation of probabilities of future records based on modeling the upper (lower) tail of the distribution for a performance measure. It uses an estimation method based on ideas from classical extreme-value theory that avoids difficulties associated with maximum likelihood method. To improve the forecasts it also utilizes other relevant information such as record times and inter-record times of the most recent records.

Application of the procedure to men's long jump and 400 meter run produces reasonable results. For long jump the last two records 8.95 and 8.90 are significantly greater that the

third best record 8.35 indicating a medium or long tail model. Data for 100 and 200- meter runs have a similar characteristic as the present records are significantly lower than the past ones.

Application of the method to Usain Bolt's individual performances prior to the 2008 Olympic shows that;

1. For him the probability of running the 200 meter in 19.31s or less was only 0.00122 indicating that his record time, 19.30, was completely unexpected.
2. The probability of beating his own best record, 19.75, was only 0.05967 indicating that his performance was exceptional.

STRENGTH-OF-SCHEDULE ADJUSTMENT TO PERFORMANCE MEASURES IN AMERICAN FOOTBALL

Jason A. Osborne[†], Chanelle Simmons

Department of Statistics, North Carolina State University

[†] E-mail: *osborne@stat.ncsu.edu*

During the Alamo Bowl in December, 2008, the announcers repeatedly piled on the pass defense of the Missouri Tigers, saying they were very nearly the worst unit in the entire country. Navigating to one of the many popular sport websites such as rivals.yahoo.com/ncaa/football/stats and choosing team defense measures, then sorting by passing yards leads to a list of 120 teams, and indeed, Missouri ranks 119. Further inspection of this list reveals several surprise members near the bottom. Also in the bottom 20, you can find five other teams from the Big 12 conference, including two of the very best teams in the country in Texas and Oklahoma, neither of which was characterized as being as bad as Missouri.

One explanation for this preponderance of Big 12 defensive units dwelling near the bottom is that they have to face more Big 12 offensive units more than other teams. The Big 12 offensive units are regarded (and ranked statistically) as being among the best in the country. The summary statistics obtained by averaging over highly imbalanced schedules are not at all sufficient. Elementary multiple regression techniques can readily be used to adjust for differences in strength of schedule to arrive at substantially different estimates of team and individual performance measures.

We investigate and compare the impact of strengh-of-schedule adjustment on performance measures in NCAA football and the National Football League, finding less impact in the NFL. We go on to control for schedule in obtaining adjusted quarterback rankings.

ARIMA MODELING OF GIANTS AND DODGERS BASEBALL SEASON STREAKS

Melvin Prince†

Southern Connecticut State University

† E-mail: *princem1@southernct.edu*

One might be disposed to use t-tests or analyses of variance to estimate comparative differences between baseball team performances over a number of seasons. Such techniques are based on the assumption that between-team and within-team variance in performance are independent between one season and the next i.e., white noise. In other words, for team performances, error variance associated with year t is presumed to be uncorrelated with that of year $t+1$. However, it is obvious that this assumption is not tenable. Winning and losing baseball team streaks from year to year, and multi-year "dynasties" are well-recognized and keenly observed by baseball fans and expert commentators.

This study's aim is to illustrate the benefit of Autoregressive Integrated Moving Average (ARIMA) modeling that allows for correlation between data in consecutive baseball seasons. ARIMA describes average levels of performance for period t in relation to measurements at p previous periods, random shocks at both period t and at the previous periods, and a between-period differencing of order d, if necessary. Giants and Dodgers year-end National League Standing between the years 1900–1957 were the data used to test for long-term and short-term comparative differences in performances between these teams. Results demonstrate differences between team performance, after taking into account series non-stationarity and ARIMA fit to specified models.

THE PYTHAGOREAN FORMULA FOR OVERTIME

Jason Rosenfeld†, Jake Fisher, Daniel Adler, Carl Morris, members of the Harvard Sports Analysis Collective

Harvard University

† E-mail: *jasonwrosenfeld@gmail.com*

Bill James, in the 1980s, created the "Pythagorean" win expectation formula, an equation that predicts a Major League Baseball team's wins based on runs scored and runs allowed. James found that previous runs scored and allowed is a better predictor of future winning

percentage than previous winning percentage because outcomes of close games are subject to luck. This formula, (runs scored)^2/ ([runs scored]^2 + [runs allowed]^2) was referred to as Pythagorean because of its resemblance to the Pythagorean formula, $a^2 + b^2 = c^2$. James' idea has been extended to other major sports, and genearlized "Pythagorean formulas," with different exponents, have emerged to predict the outcomes of games and seasons. In this paper, we estimate win expectation formulas for both regular and overtime games in the NBA, NFL, and MLB, based on games over the past 10–20 seasons. While our results for full-length games are similar to the generally accepted win expectation formulas, we believe this is the first attempt to measure how teams' runs scored and runs allowed ratios, which we term "strength," influence overtime games. We find that the impact of strength on win probability changes least in NBA overtime and most in NFL overtime. Therefore, NBA overtime games are most likely to be won by the team that would win a full-length game and NFL overtime games are most random relative to regular season games.

QUANTIFYING COACHING: EVALUATING DECISIONS OF NBA COACHES AND THEIR IMPACT ON A TEAM'S SUCCESS

Jason Rosenfeld[†], David Roher, Carl Morris, members of the Harvard Sports Analysis Collective

Harvard University

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

In recent years, statisticians have developed complex methods to measure the performance of athletes, but few useful techniques are available to quantify the contributions of managers and coaches, especially in professional basketball. The limited research that does exist is heavily contradictory. Some studies indicate that top NBA coaches can add up to 12 wins to their teams in a season, while other research shows that even Hall of Fame coaches had almost no impact on their teams' successes. Part of this gap is due to the methodology of the studies: they tend to evaluate coaching on a large-scale, season-by-season basis instead of examining the possession-by-possession coaching decisions that directly influence game outcomes.

This paper explores NBA coaches' decision-making and teams' performance through in-game metrics. It most carefully tries to answer the following questions: how intelligently does a coach time the use of his players in any given situation, holding their quality and playing time constant? Can coaches impact close games; how does a coach's team's record compare to its Pythagorean record? How well does the coach utilize tactics on the possession level, including the "2-for-1" strategy at the end of a quarter, and how does he impact the team's

performance immediately after he addresses them following a break in the action? While the paper uses basketball to illustrate its point, many of the methodologies and results in the paper are applicable to other major sports and even evaluation of other key decision makers.

MODELS FOR MOTORCYCLE GRAND PRIX RACING TIMES

Leanne Streja[†], Catherine A. Sugar, Robert E. Weiss

Department of Biostatistics, University of California Los Angeles School of Public Health

[†] E-mail: *lstreja@ucla.edu*

Motorcycle Grand Prix (MotoGP) is the most elite form of motorcycle racing with approximately 18 events per year at tracks world-wide. MotoGP teams are interested in understanding the factors predictive of fast race times in order to improve their performance/chances of winning. However, to our knowledge no one has ever directly modeled lap times and used it for strategies to win. In this analysis I generate a statistical model of lap times from MotoGP 2002-2007 race seasons with the goal of predicting results under optimal conditions. I plot and evaluate lap times throughout a race for all riders, tracks, and years of racing. I systematically remove the slow laps from the analysis to generate a model representing optimal race conditions. Several analysis of variance models are considered along with several versions of the data sets and compared to an actual MotoGP race.

THE CONTRIBUTION OF SPECIAL-TEAMS PLAY IN THE NFL

Jerry Tsai[†]

Data Domination

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

In American football, special teams refers to the set of players playing when a team intentionally relinquishes possession of the ball. Review of play-by-play data from the most recent five NFL seasons (2003-2008) reveals that on average special teams plays constitute about one-fifth of the plays in a game and account for approximately one-third of all points scored.

Rich Gosselin of the Dallas Morning News has identified 22 measures of special teams' effectiveness. Every year after the season is over, he ranks teams on each metric, and, giving each metric equal weight, constructs a summary ranking.

A normative method to measure special teams' effectiveness would be to measure the value obtained (or yielded) on each special teams play. An objective metric by which to measure this value is "Win Probability," a construct that estimates the probability of winning the game based on current context, such as the score differential, time remaining, which team has possession of the ball, down and distance, and distance to goal.

Weighting by the relative frequency each type of special team play (viz.: kickoffs and punts; field goal, extra point, and two-point conversion attempts) and role being played, the relative utility of improvements to performance on the different types of special teams plays can be estimated and indicate upon which types of plays a team should focus on improving. In similar fashion, the relative contribution of special teams to total team performance can be estimated and inform the economic efficiency of employing special team specialists.

EVALUATION OF A NEW APPLICATION OF LINEAR MODELING IN THE PREDICTION OF COLLEGE FOOTBALL BOWL OUTCOMES AND THE DEVELOPMENT OF TEAM RATINGS

Brady T. West[†1]

Institute for Social Research, University of Michigan

[†] E-mail: *bwest@umich.edu*

A 2008 paper in the Journal of Quantitative Analysis in Sports (West and Lamsal, 2008) presented a straightforward application of linear modeling in the development of a predictive model for the outcomes of college football bowl games, building on previous quantitative literature in this area. This work identified important team-level predictors of actual bowl outcomes in 2007-2008 using real Football Bowl Subdivision (FBS) data from the recently completed 2004-2006 college football seasons. Given that Bowl Championship Series (BCS) ratings are still being used to determine the teams most eligible to play for a national championship and a playoff system for determining a national champion is not yet a reality, this paper also applied the predictive model in a novel method for the calculation of ratings for selected teams, based on a round-robin playoff scenario. The objective of this presentation is to present an evaluation of the proposed method based on the results of the 2008-2009 bowl games, given that a refined predictive model was fitted using real data from the 2004-2007 seasons. Important limitations of the current method and directions for future enhancements of the method will also be discussed.

HITTING STREAKS IN SEASONS USING NON-CONSTANT BATTING AVERAGES

David Rockoff[†1], Philip Yates[2]

[1] *Iowa State University,* [2] *California State Polytechnic Pomona*

[†] E-mail: *drockoff@iastate.edu*

Many Monte Carlo simulations have been done to gauge the probability of long hitting streaks in Major League Baseball. Simulations using mean at-bats per game greatly overestimate the probability of long hitting streaks. Varying at-bats on a game-by-game basis yields more realistic results. Unfortunately, game-by-game data is largely nonexistent for seasons prior to 1953. The simulations in this presentation use a parametric model to provide an at-bat distribution for each player in each season from 1900-1953, as well as actual at-bat distributions for seasons 1954-2008. In addition, a Markov Random Field model is used to generate the probability of getting a hit in a given at-bat. Under this model, the probability is not held constant throughout the season, but rather is correlated with neighboring at-bats. This allows the simulated baseball seasons to behave as similar as possible to what goes on during the course of an actual baseball season. The simulations will identify players who have hitting streaks of at least 56 games as well as the number of simulated baseball histories that have hitting streaks of at least 56 games.