Sports Analytics in Practice – Syllabus

Stern School of Business, New York University
Professors Daniel Altman and Philip Z. Maymin
Six-week seminar – Spring 2016, Tuesdays at 6 pm

Introduction

Sports analytics is a fast-growing field that is moving far beyond the innovations in recruiting and tactics chronicled in Michael Lewis's *Moneyball*. Teams in basketball, football, ice hockey, and soccer have followed baseball in building analytics departments to support myriad aspects of decision-making on and off the field. This seminar aims to explore recent trends in sports analytics from a practical point of view, offering students the skills and ideas to create analytics of potential value to professional sports enterprises.

Registration requirements

The following requirements are non-negotiable:

- Competence in a computer language and/or software package capable of statistical analysis: R, Python, SAS, Stata, Matlab, etc.
- An understanding of basic probability, statistics, and prediction, including linear regression, properties of the normal distribution, and types of errors
- Completion of the "Firms and Markets" course

Students with questions about these requirements should contact one of the professors.

Lecturers

Daniel Altman is Adjunct Associate Professor of Economics at the NYU Stern School of Business. He is the founder of North Yard Analytics LLC, a sports data consulting firm that serves soccer teams and leagues around the world. His analysis has been featured in *The New Yorker*, The Economist, the Financial Times, The Telegraph, Bloomberg Sports StatsInsights, Quartz, and other media. He is also senior editor for economics at Foreign Policy magazine and the author of four books, including the international bestseller *Outrageous Fortunes: The Twelve Surprising Trends That Will Reshape the Global Economy*. He has a Ph.D. in Economics from Harvard University.

Philip Z. Maymin is Assistant Professor of Finance and Risk Engineering at the NYU School of Engineering. He is also the founding managing editor of Algorithmic Finance and the co-founder and co-editor-in-chief of the Journal of Sports Analytics. He has also been an analytics consultant with several NBA teams and is the Chief Data Scientist for Vantage Sports. He holds a Ph.D. in Finance from the University of Chicago, a Master's in Applied Mathematics from Harvard University, and a Bachelor's in Computer Science from Harvard University. He also holds a J.D. and is an attorney-at-law admitted to practice in California. He has been a portfolio manager at Long-Term Capital Management, Ellington Management Group, and his own hedge fund, Maymin Capital Management. He has published about two dozen research papers on behavioral and algorithmic finance and sports analytics, four books of collected essays, and a textbook, *Financial Hacking*, recently published by World Scientific.

Assignments and grading

Students will receive a large data set from a popular sport to form the basis of an analytics project of their choosing. Students may also choose to use their own data with approval of the instructors. The project will be completed within one week of the end of the course and will constitute 70% of each student's grade. The remaining 30% will come from class participation as evaluated by the professors, with knowledge of the readings as an important contributor.

Lectures

Week 1: Professor Maymin
Sports data visualization including animations and interactivity

Standard bar charts, histograms, and time series plots are often extremely unhelpful in sports contexts where the goal is usually not to present a single answer but to help refine an interesting question. Better sports data visualization does not necessarily have to be more complicated—for example, basketball shot charts are simple and effective—but it does have to do a better job at quickly providing the needed information. Human eyesight is fine-grained and attuned to pattern and complexity recognition and comparison: a good visualization is usually far more important and useful than a great quantitative model.

- **Priority Reading:** Dietrich, Carlos; Koop, David; Vo, Huy T.; Silva, Claudio T. (2014), "Baseball4D: A tool for baseball game reconstruction & visualization," 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 23-32.

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7042478&isnumber=7042473

- **Priority Reading:** Owens, Sean Gabriel; Jankun-Kelly, T.J. (2013), "Visualizations for Exploration of American Football Season and Play Data," presented at the 1st Workshop on Sports Data Visualization at IEEE VIS in Atlanta. http://workshop.sportvis.com/papers/owens_jankun-kelly_paper.pdf

- Wang, Wentao; Zhang, Jiawan; Liu, Tianshi, (2014),"MatchVis: A generalized visual multi-scale analysis framework for competitive sports," 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 287-288. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7042533&isnumber=7042473

- Maymin, Philip Z. (2013), "Acceleration in the NBA: Towards an Algorithmic Taxonomy of Basketball Plays," presented at the 2013 MIT Sloan Sports Analytics Conference in Boston. http://ssrn.com/abstract=2200560

Week 2: Professor Altman
Choosing performance metrics for players and teams

Sports executives and coaches want to measure the performance of individual players and entire teams; the resulting metrics can guide recruiting, tactics, and financial planning. But many metrics offer little or no information that can contribute meaningfully to decisions. To avoid this problem, metrics must have properties of aggregation, consistency, predictiveness, incentive compatibility, and more. These properties apply regardless of the sport in question, though in practice analysts must choose among a variety of suboptimal options.

- **Priority Reading:** Altman, Daniel (2013), "The ten traits of ideal soccer metrics." Bloomberg Sports StatsInsights (September 18) https://www.bsports.com/statsinsights/football/the-ten-traits-of-ideal-soccer-metrics#.VQcWkULWVGA

- Engelmann, Jeremias (2011), "A new player evaluation technique for players of the National Basketball Association (NBA)," presented at the 2011 MIT Sloan Sports Analytics Conference in Boston. http://www.sloansportsconference.com/wp-content/uploads/2011/08/A-New-Player-Evaluation-Technique-for-Players-of-the-National-Basketball-Association.pdf

- Pettigrew, Stephen (2015), "Assessing the offensive productivity of NHL players using in-game win probabilities," presented at the 2015 MIT Sloan Sports Analytics Conference in Boston. http://www.sloansportsconference.com/wp-content/uploads/2015/02/SSAC15-RP-Finalist-Assessing-the-offensive-productivity-of-NHL-players2.pdf

- Anderson, Chris; Sally, David (2013), "Chapter 8: O! Why a Soccer Team Is Like the Space Shuttle" in *The Numbers Game* (New York: Penguin Books)

Week 3: Professor Maymin
Big data in sports

By standard definitions, big data can mean large quantities of data, or a large variety of irregular data types, or a real-time high-frequency feed of data. In sports contexts, big data currently tends to be of much smaller breadth and scope than big data applications in other contexts, but typically too large and too irregular to fit into a standard spreadsheet. Modern sports analytics now needs to include approaches and systems for processing and analyzing optical tracking, social networks, and other comparatively large new datasets.

- **Priority Reading:** Yu, Yang; Wang, Xiao (2015), "World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets," Computers in Human Behavior, Volume 48, July 2015, pp. 392-400. http://www.sciencedirect.com/science/article/pii/S074756321500103X

- Cervone, Daniel; D'Amour, Alex; Bornn, Luke; Goldsberry, Kirk (2014), "A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes," working paper. http://arxiv.org/abs/1408.0777

- Almujahed, Sarah, et al (2013), "Sports analytics: Designing a volleyball game analysis decision-support tool using big data." Systems and Information Engineering Design Symposium (SIEDS), IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6549487 or http://catsr.ite.gmu.edu/SYST490/495_2013_SportsAnalytics/Design_of_a_Decision_Support_Tool_for_Volleyball_Using_Big_Data.pdf

Week 4: Professor Altman
Predicting outcomes of games, tournaments, and seasons

Forecasting results is important not just for bettors but also for teams, both in deciding to whether to strengthen their rosters and planning their future finances. Many forecasting models depend on power scores, but others use ratings inspired by chess rankings or simulations. Incorporating Bayesian aspects into the prediction of results is also an important innovation. A final component is the application of constraints or assumptions based on the historical behavior of teams and leagues.

- **Priority reading:** Dixon, Mark J.; Coles, Stuart G. (1997), "Modelling Association Football Scores and Inefficiencies in the Football Betting Market," *Applied Statistics*, Volume 46, Number 2, pp. 265-280 http://onlinelibrary.wiley.com/doi/10.1111/1467-9876.00065/abstract

- Hvattum, Lars Magnus; Arntzen, Halvard (2010), "Using ELO ratings for match result prediction in association football," International Journal of Forecasting, 2010, Volume 26, Number 3, pp. 460-470 http://www.sciencedirect.com/science/article/pii/S0169207009001708

- Constantinou, Anthony C.; Fenton, Norman E.; Neil, Martin (2012), "pi-football: A Bayesian network model for forecasting Association Football match outcomes," *Knowledge-Based Systems*, Volume 36 (December), pp. 322-339. http://www.sciencedirect.com/science/article/pii/S0950705112001967

Week 5: Professor Maymin
Machine learning in sports

The original statistical revolution in sports was largely an inspired application of linear regression to available data. The next wave will come from leveraging the advances in machine learning techniques, allowing a computer to learn on its own from past data. Some of the hottest techniques include random forests, support vector machines, nearest neighbors, and deep learning. We will go over the basics of these approaches, and look at applications to drafts and player projections from past performance.

- **Priority Reading:** Maymin, Philip Z. (2014), "The Automated General Manager: An Unbiased, Backtested Algorithmic System for Drafts, Trades, and Free Agency that Outperforms Human Front Offices," working paper.

- Lee, Greg; Bulitko, Vadim; Ludvig, Elliot (2014), "Automated story selection for color commentary in sports," Computational Intelligence and AI in Games, IEEE Transactions on 6.2: 144-155.

- Kyriakides, George; Talattinis, Kyriacos; George, Stephanides (2014), "Rating Systems Vs Machine Learning on the context of sports," in Proceedings of the 18th Panhellenic Conference on Informatics (PCI '14). ACM, New York, NY, USA.

Week 6: Professor Altman
Evaluating performance through profit and loss

Sports analysts frequently assume that the teams they are modeling have points, wins, positions in league standings, or trophies as their objectives. But sports teams are businesses, and sometimes even public companies. As a result, the ultimate objective of many decisions may be long-term profitability. Transforming game-based metrics into cash-based metrics is therefore an essential part of the analyst's toolkit. We will consider how to evaluate players as capital assets and assess their transfer or trade values.

- **Priority Reading:** Altman, Daniel (2015), "Player contributions to cash flow," presentation. http://www.northyardanalytics.com/NYA-player-contributions.pdf

- **Priority Reading:** Altman, Daniel (2015), "Managing the player portfolio," presentation http://www.northyardanalytics.com/NYA-managing-player-portfolio.pdf

- Baumer, Benjamin; Zimbalist, Andrew. (2014) "Chapter 6: Analytics and the Business of Baseball" in The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball (Philadelphia: University of Pennsylvania Press)