

**BOSTON COLLEGE**  
Department of Economics

**EC 370.01: Sports Econometrics (Spring 2018)**  
Stokes S295: TTh (1:30 – 2:45)

Christopher Maxwell  
[maxwellc@bc.edu](mailto:maxwellc@bc.edu)  
<http://www.cmaxxsports.com>

Maloney Hall, 337  
Hrs: TTh 3+  
... & by arrangement

**Course Description:** This is an advanced stats/econometrics course; it is not a sports history or trivia class. We'll be developing various statistical tools of analysis and then applying those tools to a wide variety of sport-related topics, perhaps including:

- rating teams and forecasting team performance,
- predicting game outcomes,
- the efficiency of wagering markets,
- the drivers of home field advantage in sports,
- ... and umpire/referee bias in sports,
- the business and economics of professional team sports,
- measuring and valuing parity in sports leagues,
- the importance of population in driving competitive imbalance,
- the efficacy of leagues' competitive balance initiatives,
- peer effects in team performance,
- the relationship between performance and player compensation,
- understanding the what drives ticket prices,
- valuing draft picks,
- ... and so forth.

We could easily work with other data, but there's so much publicly available sports-related data available... so why not? ... and besides, it's so much fun!

We will also weave some sports economics into the course. Most of that material will focus on the notion of *competitive balance* and the *Uncertainty of Outcome Hypothesis*, which many would say is the most important concept in sports economics.

And again: ***This is not a sports history or trivia class.***

**Prerequisites:** Intermediate Microeconomics (EC201 or EC203) and Econometrics (EC228 and/or EC327). Students are expected to have excelled in EC228, and to know how to run basic econometric models (OLS: SLR and MLR) using Stata ... and to be comfortable interpreting regression results.

This course will make extensive use of both Excel and Stata:

- You should have worked with Stata in your Econometrics course. At the start of the semester, we will review how to access and run Stata through BC's apps server.

## EC 370: Sports Econometrics

To avoid traffic jams with Citrix and the apps server, you may want to purchase a six-month Stata IC license for \$45 (sorry, but small Stata will not suffice for EC 370). For details, go to: <http://www.stata.com/order/new/edu/gradplans/student-pricing/>.

- This course also makes extensive use of Excel. You should not take this course if you do not have strong Excel skills. To brush up on your Excel skills, you might take a look at <https://www.bc.edu/offices/help/teaching/elearning/elearningcourses.html> (Microsoft's e-Learning courses).

Unfortunately, the features offered by Excel differ somewhat across platforms and over time. For this course, you may need to install Excel's *Analysis ToolPak* and *SolverAdd-in*. If your version of Excel is anything other than Excel 2011 for the Mac, this should be straightforward. If you have Excel 2011 for the Mac, then you have at least two options: 1) run Excel on BC's apps server, which has these capabilities and is faster than you might imagine, or 2) go to <http://www.bc.edu/software/applications/office.html> and download and install Excel 2016.

**Analytic tools/methods:** While the list may change, at the moment I anticipate that we'll be focusing on the following statistical tools of analysis, working with both Stata and Excel:

- SLR and MLR estimation and inference (review)
- Assessing importance/meaningfulness of estimates: elasticities and beta regressions
- Non-linear least squares (and the *residual*)
- Binary dependent variables – I: Linear (and truncated linear) probability models (LPMs)
- Functional forms: fixed effects, percentile dummies, polynomials, and splines (linear and cubic)
- Binary dependent variables - II Maximum likelihood estimation (MLE), and logit, probit and even arc-tangent models
- Runs, streaks and testing independence: chi squared tests, binomial tests, regression analysis and *runs* tests
- More about limited dependent variables: Ordered logit and probit, multinomial logit, and poisson and negative binomial count models

**Applications:** We'll illustrate the analytic methods with multiple applications, working with sports-related data... often working in both Excel and Stata. The applications that I have in mind fall into the following general categories:

- Strategy and valuing game-states
- Umpire/referee/judging bias (and home field advantage)
- Forecasting game/match outcomes (e.g. runs, wins, losses, ties, points, putts, etc.)
- Assessing player and team performance
- Efficiency of wagering markets
- Pay and performance (for players, coaches and teams)
- Momentum effects (streaks and runs)

## EC 370: Sports Econometrics

### Texts:

- **Required:** Tobias Moskowitz and Jon Wertheim, **Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won**, Three Rivers Press (paperback), 2012.
- **Recommended, but not required:** Rodney Fort, **Sports Economics**, Prentice Hall.

**Canvas:** All course related material will be posted to Canvas.

**Accommodations:** If you are a student with a documented disability seeking reasonable accommodations in this course, please contact Kathy Duggan (x2-8093; dugganka@bc.edu) at the Connors Family Learning Center regarding learning disabilities and ADHD, or Paulette Durrett, (x2-3470; paulette.durrett@bc.edu) in the Disability Services Office regarding all other types of disabilities, including temporary disabilities. Advance notice and appropriate documentation are required for accommodations.

**Academic Integrity:** You will be held to Boston College's standards of academic integrity. If you have any questions as to what that means, please go to <http://www.bc.edu/offices/stserv/academic/integrity.html>.

**Pass/Fail:** It's perfectly OK to take this course Pass/Fail; however it is not fair to your fellow students to shirk on team assignments. I expect equal contributions by all team members, but history has proved that students taking the course Pass/Fail have at times unfortunately failed to pull their weight. And that's just not fair to the other students. If you are taking the course Pass/Fail, please let me know at the start of the semester.

**Course Structure:** There are four graded elements in the course; they are (%'s of course grade are in parentheses):

1. One mid-term exam (35%)
2. Six-or-so exercises (35%)
3. Empirical research project and presentations (25%)
4. Tuesday topics/Participation (5%)

### 1. Mid-Term Exam (35% of total grade):

There is one exam in this course... a Mid-Term exam covering the empirical methods and applications developed in this course. Exam grades are curved. The exam date is yet to be determined, but I'm thinking about some time in the next to last week of classes, perhaps Thursday, April 26<sup>th</sup>.

Note: Only in extraordinarily compelling situations will I even consider the possibility of a "make up" exam. It is your responsibility to plan your schedule accordingly.

### 2. Exercises (35% of total grade):

I anticipate having six-or-so exercises (of equal value) over the course of the semester. These will typically be team assignments (usually with two students per team) lasting about two weeks.

## EC 370: Sports Econometrics

I will assign the teams, which will change from exercise to exercise. The set of exercises is not yet set, but here's what I have in mind at the moment:<sup>1</sup>

- a. *Moneyball* Revisited: OBP v. SLG in MLB (review of SLR and MLR analysis; measuring explanatory power; statistical significance v. importance/meaningfulness (economic significance); elasticities; *beta* regressions)
- b. Valuing Draft Picks using NFL Trade Data (Weibull and exponential distributions; non-linear least-squares estimation; the importance (and at times, arbitrariness) of the residual)
- c. *March Madness* and the Ratings Performance Index (RPI) in NCAA Basketball (ratings models; model calibration; concordance; Kendall *tau* correlations; non-linear estimation)
- d. Valuing Gamestates: Run Production in MLB (*quick slant*; fixed effects; 14+ million observations)
- e. Umpire/Referee Bias and Home Field Advantage in MLB (binary dependent variables; linear probability models; MLE estimation (logit and probit models); estimating bias; categorical dummies; interaction effects)
- f. *Topping the Table*: The 2015-16 Leicester City Foxes shocked the world (Premier League football) (trivariate dependent variables (win/draw/lose); ordered logit/probit; multinomial logit/probit; poisson and negative binomial count models; bivariate poisson with correlated errors)

In many cases, there are faster and slower ways to complete the exercises. Let me know if progress is painfully slow, and I'll be happy to make suggestions to help speed things up. No late work accepted. Final grades on Exercises are curved.

### 3. Empirical Research Project (25% of total grade):

The empirical research project will kick off with team assignments in the week after Spring Break. I will assign project teams, which will likely have two or three members each. Students' grades will reflect both their individual performance as well as the quality of the final team product.

***Shirkers take notice:*** *Peer evaluation* forms will be distributed at the end of the semester, so that team members can assess each other's performance.

The final deliverable is a PowerPoint presentation (or equivalent). You can structure these presentations as you wish, but I suggest the following six sections at a minimum:

1. Introduction (description of topic and summary of results)
2. Brief literature review
3. Description of model and nature of analysis
4. Discussion of data
5. Presentation of results
6. Conclusion

---

<sup>1</sup> I have prepared a fairly large array (35 and counting!) of exercises over the course of the 10+ years teaching this course.... so there's plenty to choose from! These five are favorites... but there are many others, that we might do.

## EC 370: Sports Econometrics

In all cases please be concise and to the point; *shorter is always better*. I will say more about the format of the deliverable when teams are assigned. *Empirical work is slow going. Be sure to leave yourself enough time to complete the assignment to your satisfaction.*

There are three or four milestone dates... depending on how you count:

### Phase I: *Topic selection* (Thurs. April 5<sup>th</sup>)

There are two deliverables due on the 5<sup>th</sup>:

- A PowerPoint presentation, and an
- In-class presentation

At a minimum, you should 1) present your topic, 2) review in detail at least two related papers of interest, and 3) briefly discuss the data that you'll be working with.

Please send me a one paragraph description of your topic by 6 PM on 4/2 (so I have time to put them into a handout for the class).

The *Topic Selection* PowerPoint deliverable will be graded Pass/Fail. A failing grade will require revisions and resubmissions until a passing grade is achieved.

### Phase II: *Progress report* (Thurs. April 19<sup>th</sup> ... so as to avoid *Marathon Tuesday*)

As in Phase I, there are two deliverables due on the 19<sup>th</sup>:

- A PowerPoint presentation, and an
- In-class presentation

In your presentations you should discuss 1) your completed literature review, 2) a status report on your data work, including summary stats and a detailed discussion of the data, and 3) a discussion of next steps and *todos*.

### Phase III: *Research project presentations* (Thurs. May 3<sup>rd</sup> ... last class)

Again, the same two deliverables:

- A PowerPoint presentation, and an
- In-class presentation

This is your chance to brag about your amazing work. Please send me a one paragraph description of your research project by 6 PM on 5/2 (so I have time to put them into a handout). Your paragraph should include 1) your topic, 2) what you did, and 3) what you found.

If you feel good about your work, you can submit your final draft presentation on this date. Alternatively, you can incorporate some of the feedback that you receive in your presentation, and revise/update your presentation accordingly.

**Final drafts:** due by 5PM on Monday, May 7<sup>th</sup>.

Final drafts are due in hardcopy form in my mailbox in the Economics Dept. mailroom.

Just to be clear, there are three/four milestone dates/deliverables:

1. **April 3<sup>rd</sup>** - Topic selection: topic; two related papers/studies; data sources
2. **April 19<sup>th</sup>** - Progress report: Literature survey; data discussion; outlook

## EC 370: Sports Econometrics

3. **May 3<sup>rd</sup>** - Research project presentations (final draft might be submitted at this time)
4. **May 7<sup>th</sup> 5PM** - Final draft final due date

### 4. Tuesday Topics (5% of total grade)

These will typically take place at the start of every Tuesday class (if we need more slots, we'll add some Thursday presentations). We'll devote the first 10 minutes or so of class time to a discussion of a current relevant issue. Given the class size, the discussion will be led by a team of three students (team assignments will be distributed once the class list is finalized). The team leading the discussion may want to prepare a PowerPoint presentation to guide and focus the discussion. I hope that your presentations will include some of **your own** empirical analysis of the topic. To provide a sense of how this might work, I'll do the first presentation. Presentations will be graded, and along with *participation*, count towards 5% of your course grade.

**Important:** You will be limited to 10-15 minutes, and at most six slides (not counting the title slide).

**Proposed Schedule of Topics:** The schedule and set of applications will likely evolve as we work through the semester, but here's a sense of the topic schedule:

### *Econometrics and (other) Statistics*

1. Review of Simple Linear Regression (SLR) analysis (Excel & Stata)
  - a. MLB: The Pythagorean Theorem; NBA: shooting success and distance; NFL: field goals success and distance
2. Review of Multiple Linear Regression (MLR) analysis (Excel & Stata)
  - a. MLB: The Pythagorean Theorem; NFL: Ticket Prices; NBA referees and *own-race* bias; NFL: field goal success, altitude and distance
3. Non Linear regression analysis ... and the *residual* (Excel & Stata)
  - a. MLB: The Pythagorean Theorem; NFL: field goal success, altitude and distance
4. Binary Dependent Variables I: The *Linear Probability Model* (LPM) (Excel & Stata)... we'll look at both simple LPMs as well as *truncated* LPMs
  - a. NBA: *The Myth of the Hot Hand*; NFL: Icing the Kicker; MLB: *Win Expectancy* and *Leverage*; PGA: *Putting Prowess*; NFL: *Deflategate*
5. Binary Dependent Variables II: Maximum Likelihood Estimation (MLE) (Excel & Stata)... logit, probit and even arc-tangent
  - a. Repeat the examples in 4a.
6. Functional forms (percentile dummies, polynomials, splines (linear and cubic) and fixed effects; big v. small datasets) (Stata)
  - a. NBA: shooting success; NFL: field goals tries; PGA: *Putting Prowess*

## EC 370: Sports Econometrics

7. Testing independence in play calling and success: Chi-squared tests, Binomial tests, OLS analysis and the *Wald Wolfowitz* runs test (Excel & Stata)
  - a. NBA: More Hot Hand; MLB: pitch selection; maybe PGA or tennis?

**Selected Topics:** The topic list may very well change depending on time and interest.

8. Competitive balance and the *Uncertainty of Outcome* Hypothesis
  - a. It's only the most important concept in sports economics, so we might as well spend some time on it:)
  - b. Theory and evidence (MLB, NBA, NFL, NHL and European football/soccer) (at the season and game levels, and across seasons); testing the *Uncertainty of Outcome* Hypothesis
9. *Retrodictive* and *predictive* ratings models: Assessing team performance; forecasting game/match outcomes
  - a. Simple wins and points models – OLS, LPMs, logit and probit (NCAA football); ordered logit and probit models (European football); the *google PageRank* model (NCAA basketball)
  - b. MLE Count models of goals/points – multinomial logit, poisson and negative binomial regressions, correlated errors (European football)
10. Wagering market efficiency
  - a. Football (NCAA and NFL (lines and spreads)); Thoroughbred racing (pari-mutuel odds); European football (lots of odds data!)
11. Peer effects in performance
  - a. Estimating production complementarities (NBA synergies on the court; NHL too?)

### Additional Resources

- Rodney Fort: <https://sites.google.com/site/rodswebpages/codes>
- John Vrooman: <https://my.vanderbilt.edu/vrooman/>
- Journal of Sports Economics (JSE): <http://jse.sagepub.com/>
- Journal of Quantitative Analysis in Sports (JQAS): <http://www.degruyter.com/view/j/jqas>
- Multi-author blog: [www.thesportseconomist.com](http://www.thesportseconomist.com)
- Sports Business Daily: <http://www.sportsbusinessdaily.com/Daily.aspx> (expensive but informative; two week trial subscription; student rates (still expensive))
- Sports Business Journal: <http://www.sportsbusinessdaily.com/Journal.aspx> (I believe the library has acquired a subscription to this journal)
- SportsBiz: <http://thesportsbizblog.blogspot.com/>
- Sports Law: <http://sports-law.blogspot.com/>
- National Sports Law Institute (Marquette): <http://law.marquette.edu/national-sports-law-institute/welcome>
- The “Wages of Wins” Journal: <http://dberri.wordpress.com/>
- and <http://www.cmaxxsports.com/misc/resources.html> (you'll find useful web pages devoted to MLB, the NBA, the NCAA, the NFL, and European football/soccer... and more)