Honors Thesis Proposal

I. Project Name:

Predicting Regular Season Results of NBA Teams Based on Time Series and Regression Analysis of Common Basketball Statistics

II. Project Background:

As a fanatic basketball fan, I watched NBA games online, on TV, in the bars and in the arenas countless times. Since I started college at Berkeley and went to a few Warriors' home games, I quickly became a Warriors' fan. I have been absolutely enjoying watching Stephen Curry and Klay Thompson putting on a show night in and night out. Fortunately, having been a solid team for several consecutive seasons, the Warriors finally unleashed their full potential this season and they currently have the best record in the NBA, breaking numerous franchise records along the way.

Even though basketball teams only play 82 regular season games, not as many as 162 games baseball teams play during a season, there still is significant amount of data generated from all these games. I have seen a number of papers analyzing baseball teams based on their players' batting averages and other common statistics, but very few researches have touched upon other sports like basketball.

Boxscores tell the story of each game, and I have been reading them religiously every night. Some sports website and analysts have also come up with creative metrics such as Player Efficiency Index to gauge the effectiveness of individual players, but few has taken a step further to correlate players' individual statistics with a team's success and make "scientific" predictions of teams' performances.

I believe with the abundance of data available through sports website (such as ESPN and Yahoo Sports where they provide APIs for people to access their database) and NBA.com, I would be able to come up with some metrics based on common statistics to evaluate a team's strength (and weakness). Equipped with statistical tools that I learned from previous statistics classes (such as Stat 133, 134, 135, 150 and 153), I hope to

explore the correlation between individual players' statistics and their teams' success. Time series and regression analysis could be especially helpful. My ultimate goal is to **make predictions about teams' regular season results** (I refrain from making predictions about the NBA championship because Playoffs generally have very different dynamics than regular season games, and more factors come into play such as match-ups, injuries, etc., and the results may not be as consistent as regular season games).

III. Data to Collect:

I plan to collect my data through sports networks (such as ESPN and NBA.com) for the past ten to twenty seasons. The statistics that I plan to use include individual players' TS, AST, TO, USG, ORR, DRR, REBR, PER, VA (please see section VII for Glossary), among other statistics. I will also collect all NBA teams' data such as their win-loss percentages, team scoring, rebounding, assists, turnovers stats, etc. during the same period as the individual statistics. After acquiring those data, I would explore the correlation between a team's record and its players' individual statistics using statistical methods and tools. I would be primarily using R to process and manipulate data.

IV. Questions to Address:

As mentioned earlier, the ultimate goal is to predict the regular season results of the NBA teams. Now that we are half way through the 2014-15 season, it is likely that I can find out whether my findings and predictions make sense before the semester ends, which is quite exciting. I would use historical data to test my theories before using them to make predictions.

V. Relevant Statistical Methods:

Not entirely sure about what specific techniques that I will be using, but I think **stationarity analysis, trend / model fitting**, **regression analysis**, among others, are likely to be helpful to analyze the data set. I will also try to conduct **spectral analysis** to see if there is any recurring trend.

VI. References to Pertinent Work:

I will probably make references to some of the work done relating to baseball statistics. Furthermore, I would test, refer to, and incorporate some metrics developed by sports analysts, such as Hollinger's NBA Player / Team Statistics in my analysis. I will certainly do more research into this area and make reference to other relevant work should this proposal be approved.

VII. Glossary[1]:

TS%: True Shooting Percentage - what a player's shooting percentage would be if we accounted for free throws and 3-pointers. True Shooting Percentage = Total points / [(FGA + (0.44 x FTA)]

AST: Assist Ratio - the percentage of a player's possessions that ends in an assist. Assist Ratio = (Assists x 100) divided by [(FGA + (FTA x 0.44) + Assists + Turnovers]

TO: Turnover Ratio - the percentage of a player's possessions that end in a turnover. Turnover Ratio = (Turnover x 100) divided by [(FGA + (FTA x 0.44) + Assists + Turnovers]

USG: Usage Rate - the number of possessions a player uses per 40 minutes. Usage Rate = {[FGA + (FT Att. x 0.44) + (Ast x 0.33) + TO] x 40 x League Pace} divided by (Minutes x Team Pace)

ORR: Offensive rebound rate

DRR: Defensive rebound rate

REBR: Rebound Rate - the percentage of missed shots that a player rebounds. Rebound Rate = (100 x (Rebounds x Team Minutes)) divided by [Player Minutes x (Team Rebounds + Opponent Rebounds)]

PER: Player Efficiency Rating is the overall rating of a player's per-minute statistical production

VA: Value Added - the estimated number of points a player adds to a team's season total above what a 'replacement player' (for instance, the 12th man on the roster) would produce. Value Added = ([Minutes * (PER - PRL)] / 67). PRL (Position Replacement Level) = 11.5 for power forwards, 11.0 for point guards, 10.6 for centers, 10.5 for shooting guards and small forwards

[1] 2014-15 Hollinger NBA Player Statistics, http://insider.espn.go.com/nba/hollinger/statistics