How Much Trouble Is Early Foul Trouble?

Allan Maymin (AllianceBernstein), Philip Maymin (NYU-Poly), Eugene Shen (AllianceBernstein)
NYU-Polytechnic Institute
Six MetroTech Center, Brooklyn, NY 11201
Email: phil@maymin.com

Abstract

We introduce a win probability approach to modeling basketball performance and employ it to determine the effect that early foul trouble by a team's starters has on its future performance. We find that each of the three seasons from 2006-2007 through 2008-2009 display negative team performance when starters are in early foul trouble, defined as having committed at least one more foul than the current quarter ("Q+1"). Most players in early foul trouble should be yanked until they are no longer in foul trouble. Our approach can be extended to other state variables.

1 Introduction

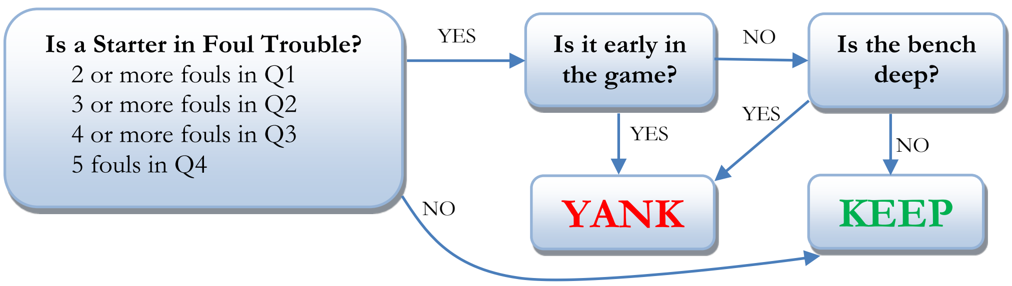
Should starters in early foul trouble be yanked? The advantage of yanking is that the starter will likely be able to play at the crucial end of the game but the disadvantage is that he may not play as many minutes as he otherwise would. On the other hand, if a starter is kept in the game, he may not play at his full potential, as the opposing team tries to induce him to commit another foul. Thus, the coach must evaluate these tradeoffs. Different coaches may have different definitions of how many fouls constitute "trouble," and whether it is optimal to yank or not. Our research shows that it is optimal to yank starters on a "Q+1" basis, i.e., when they commit one more foul than the current quarter. Indeed, the effect is so strong that a single incorrect decision could decide the game.

Our research shows that players in foul trouble play worse than they normally do. For example, in Game 2 of the 2010 NBA Finals between the Boston Celtics and the LA Lakers, Kobe Bryant picked up his fifth foul with 11:15 remaining in the fourth quarter. Phil Jackson removed Bryant from the game at the 8:11 mark, and reinserted him in the game with 6:16 remaining and the game tied at 85-85. After he re-entered the game, Bryant played tentatively. Steve Aschburner of NBA.com [1] noted that "[Kobe's] desire to stick around can cause him to back off ever so slightly, avoid a defensive encounter, drop his attack gear. Play c-a-r-e-f-u-l-l-y. And self-consciously, in a way that can distract even an assassin from his task." J.J. Adande of ESPN [2] observed that Kobe "couldn't afford to be aggressive. It was like watching a pitcher who didn't have his fastball rely on mixing speeds and changing locations. Bryant took a little longer to initiate his moves, surveying the defense to determine where the late help would come from. He tried jab-stepping a couple of times to see if he could force their hand, the equivalent of a football team sending a receiver in motion. When he did go on the attack, he had to pull up early, leading to some amazing examples of body control (and yet another head-scratching call, on Kendrick Perkins, for an and-one after Bryant specifically stopped short of him to avoid contact)." The Celtics won the game 103-94, outscoring the Lakers 18-9 in the final six minutes. After the game, Kobe conceded [2] that when playing with foul trouble, "You've just got to be careful."

We find that Kobe's tentative play is not unusual. We use a large dataset of play-by-play NBA data to determine that teams play worse if they play foul-plagued starters. We use a novel win-probability technique to measure the impact of foul trouble on a team's probability of winning. This win-probability technique is sufficiently general to be useful for other questions simply by appropriately redefining the state variables. Thus, our two contributions are to introduce the new approach and to apply it to the problem of early foul trouble that had remained unaddressed in the academic literature.

We believe that our research has valuable insights for basketball coaches. We find that coaches in general should yank starters who are plagued with foul trouble, with foul trouble defined by the "Q+1" rule. There are, however, nuances that need to be considered. Coaches must also consider the time remaining in the game, and the quality of the player with foul trouble. Early in the game, yanking a player preserves "option value" since the coach can reinsert a fresh, non-foul plagued starter back into the game in the fourth quarter. Thus, yanking makes more sense early in the game. Also, the coach must evaluate the quality of the player relative to his potential substitute. If the starter is not sufficiently better than his substitute, then he ought to be yanked from the game to preserve his option value. On the other hand, a superstar in foul trouble (e.g. Kobe Bryant) may still be better than a fresh bench player. In this case, the coaches may be better off keeping the starter in the game.

Should a Starter in Foul Trouble be Benched? A Flowchart



2 Description of Data

We examine play-by-play data from the NBA for the 2006-2007, 2007-2008, and 2008-2009 seasons. This data is from the website `http://basketballgeek.com/data`, maintained by Ryan J. Parker, and represents a processed version of the play-by-play information from the NBA and ESPN. The data includes such information as the names of all players on the court at each time, the location of the shots taken, the reasons for fouls called, and more. For our purposes, we use the players on the court and the substitutions, as well as the fouls.[1] There are 1,149 games in 2006-2007, 1,184 in 2007-2008,

[1] The running score of the game is not provided in 2006-2007 though it is in later seasons. For 2006-2007, we manually calculate the running score by combining free throws and field goals, with adjustments for three-point field goals. These score calculations occasionally differ from the official final score of the game, for example if a three point field goal had been incorrectly ruled a two point field goal during the course of the game. To counteract this discrepancy, we fix the final score of each game to be the official final score, and adjust the initial net score by the difference between our calculated final net score

and 1,176 in 2008-2009 with sufficient data available.[2] We use nearly every tick in the database, eliminating only those that do not lead to a change in a state variable.

We distinguish "threshold" fouls, namely those that move a player into foul trouble, from other fouls with the commonly-used "Q+1" measure (i.e., yanking players who commit their third foul in the second quarter, or fourth foul in the third quarter, etc.). Relative to other measures such as those that depend on the ratio of the remaining fouls to the remaining game time, "Q+1" is the best match to actual coaching decisions. For a comparable number of threshold fouls, "Q+1" fouls had a much higher percentage of yanks, indicating that it is the more common rule that coaches follow.

A player is "yanked" by his coach if he is substituted out of the game within some number of seconds of committing a foul. We use 30 seconds to allow for coaches to keep their player in for one more offensive possession, where they may be less likely to commit a foul and more likely to help their team relative to the expected substitution.

Finally, we define the following variables:

- STA: net number of starters in the game
- FTR: net number of starters in the game with foul trouble, where foul trouble is defined by the "Q+1" criteria
- cFTR: net sum of starters' "continuous FTR", where a continuous FTR is defined as how many quarters must elapse before the player is no longer in "Q+1" foul trouble. For example, if a player picks up his third foul at the beginning of the second quarter, then his continuous FTR is 1, while if he picks up his third foul halfway through the second quarter, his continuous FTR is 0.5.

These variables are net (home minus away).

3 Model Specification: Win Probability Framework

Our specification can be compared with Rosenbaum's [3] advanced plus/minus model or Moskowitz and Wetheim's [5] estimates of player effectiveness conditional on foul trouble: rather than using players, we use teams, and rather than estimating the impact on the subsequent point differential, we measure the impact on a team's probability of winning. These differences allow us to incorporate state variables and variance by extending a general statistical model of team performance as described in Stern [4]. Furthermore, by focusing on the probability of winning, we employ a "macro"-level analysis that could complement or be complemented with the "micro"-level analysis of a plus/minus-type model.

---

and the official final net score. We do this for consistency of scores within a game, but our results do not differ materially if we use calculated score only. To maintain consistency, we do this for later years as well.

[2] Play-by-play transcripts for games occasionally fail to include sufficient information to determine which players are on the court, and so such games are ignored. Such errors by the NBA and ESPN seem to be random.

Stern models the probability of the home team winning as $\mathcal{N}(F_t/\sigma_t)$ where $\mathcal{N}(\cdot)$ is the cdf of the standard normal distribution, $0 \leq t \leq 1$ is the fraction of the game elapsed, the variance σ_t^2 is proportional to $1-t$ such that $\sigma_t^2 = (1-t)\sigma^2$ for some volatility constant σ, and F_t is the "forward lead" at time t:

$$F_t = \beta_\ell \ell_t + (1-t)\mu$$

where ℓ_t is the current lead, β_ℓ is the coefficient for the lead, and μ is the drift constant.

Our "Win Probability" framework uses Stern's model with a few extensions. Our first and most general innovation is to add other state information to the definition of the forward lead. This is a better approach than simply regressing point differentials on various variables because it automatically dampens (heightens) the effects of early (late) game actions when the variance is greater (lower). Further, it highlights the importance of the information ratio in determining net benefit or cost, thus taking into account the risk as well as the return. Our second innovation is to model variance as declining linearly with time at a rate lower than one. These innovations are described in detail below.

First, we add possession, team dummies, a constant, starter and foul variables to Stern's definition of forward lead:

$$F_t = \alpha + \beta_\ell \ell_t + \beta_P P_t + (1-t)\left(\mu + \sum_{i=1}^{29} \beta_i D_{it} + \beta_{STA} STA_t + \beta_{FTR} FTR_t\right)$$

where α is a constant, P_t represents possession at time t and β_P is its coefficient, β_{STA} and β_{FTR} are the coefficients associated with STA and FTR, and the D_{it} are the 29 team dummy variables with β_i as their associated coefficient. The coefficient of the 30th team dummy (the Washington Wizards) is normalized to zero. We multiply STA and FTR by the fraction of the game remaining because these variables affect the rate of change of the lead.

When the game begins, each team has 5 starters without foul trouble:

STA(home) = STA(away) = 5
FTR(home) = FTR(away) = 0

And of course the initial net numbers are both zero:

STA = STA(home) – STA(away) = 0
FTR = FTR(home) – FTR(away) = 0

If the home team has a starter in foul trouble, then the coach faces the following choice: either keep the starter in the game (hence, increment FTR), or yank the starter from the game (hence, decrement STA but leave FTR unchanged).

We expect β_{STA} to be positive, since teams play better if their starters are in the game. The optimality of yanking will depend on the size of β_{FTR}. If $\beta_{STA} + \beta_{FTR} < 0$, then it will be optimal to yank, since the starter in foul trouble plays worse than a bench player. If $\beta_{STA} + \beta_{FTR} > 0$, then it may be suboptimal to yank locally, but it may still be optimal to yank globally since yanking preserves option value by reducing the probability that the player remains in foul trouble.

Next, we model variance as a first order function of time: $\sigma_t^2 = 1 - \gamma t$, where γ (gamma) will be estimated below. Volatility at time $t = 0$ is normalized to be 1.

This volatility specification is analogous to a "jump-diffusion" model in finance such as Merton [6], where the constant represents a jump while the gamma represents diffusion. This specification seems appropriate for basketball. Variance should be proportional to possessions, which generally increases linearly with time remaining. At the end of a game, however, there is more variance than would be implied by a pure diffusion model, since teams can squeeze many possessions into the final minute of the game with fouls and time-outs.

4 Key Research Findings

4.1. Summary Statistics and Descriptive Measures

Figure 1 graphs the total number of threshold fouls and yanks for each team during the 2006-2007 NBA season. The Golden State Warriors are the biggest outlier by two measures: they had the most threshold fouls, so they are further to the right on the graph, and they had the fewest yanks per threshold foul, so they have the lowest slope (the Washington Wizards had the second lowest slope, and the average slope was 0.69).

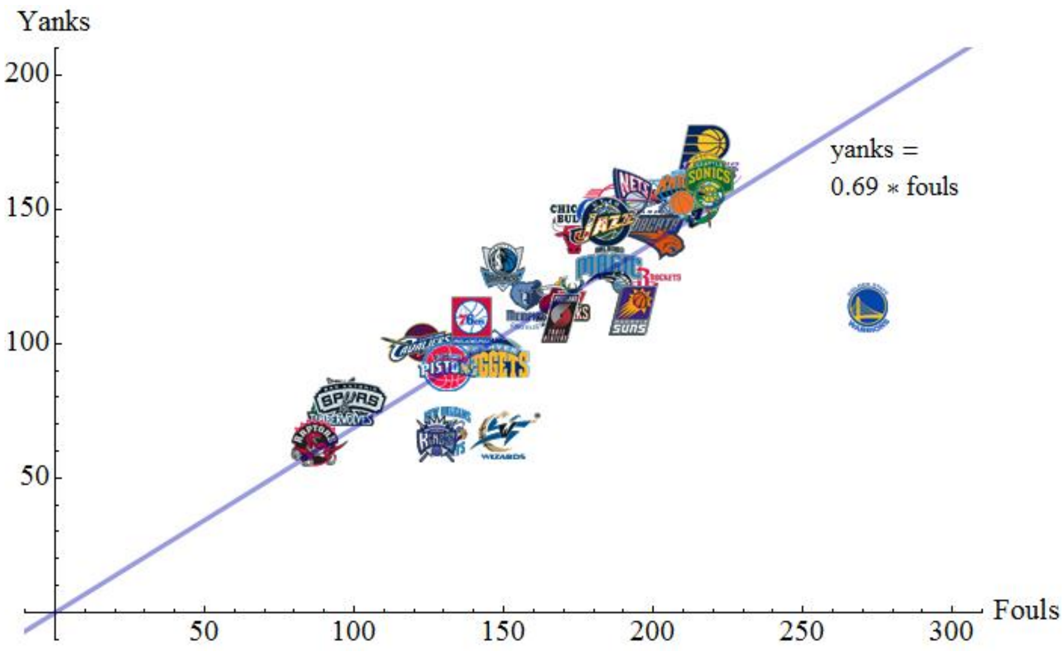


Figure 1: Total Threshold Fouls and Yanks per Team in the 2006-2007 season

Table 1 lists the top 10 in various categories. The first category is the total number of fouls committed, led by Samuel Dalembert with 280. The second category is the total number of threshold "Q+1" fouls committed. The third category is percentage of time that a player was yanked by his coach immediately after committing a threshold "Q+1" foul. The fourth category is the percentage of time that a player was not yanked in that situation.

Jeff Foster was yanked every single time he committed a threshold foul; Baron Davis was kept in the game 68 percent of the time he committed a threshold foul. The league average was 15 threshold fouls per player. The ranking methodology only includes those players that committed an above-average number of threshold fouls, to ensure sufficient data observations for a coach's yank/keep tendency to be revealed.

Table 1: Leaders in Fouls and Threshold Yanks and Non-Yanks for 2006-2007

Rank	Total Fouls	Threshold Fouls	Yanks Per Threshold Foul	Non-Yanks (Keeps) Per Threshold Foul
1	Samuel Dalembert (280)	Jason Collins (71)	Johan Petro (100%)	Baron Davis (68%)
2	Amare Stoudemire (266)	Chuck Hayes (70)	Quinton Ross (100%)	Chris Paul (53%)
3	Jason Collins (261)	Amare Stoudemire (70)	Jeff Foster (100%)	Jason Richardson (50%)
4	Al Harrington (258)	Andris Biedrins (67)	Ben Gordon (96%)	Mickael Pietrus (50%)
5	Kirk Hinrich (254)	Nick Collison (66)	Kendrick Perkins (95%)	DeShawn Stevenson (47%)
6	Zydrunas Ilgauskas (249)	Kirk Hinrich (64)	Jarrett Jack (95%)	Monta Ellis (45%)
7	Gilbert Arenas (249)	Erick Dampier (62)	Fabricio Oberto (95%)	Stephen Jackson (45%)
8	Vince Carter (246)	Samuel Dalembert (60)	Mike Miller (95%)	Gilbert Arenas (42%)
9	Chris Wilcox (245)	Chris Kaman (59)	Shelden Williams (95%)	Chuck Hayes (41%)
10	Dwight Howard (242)	Andrew Bynum (57)	Luol Deng (94%)	Raja Bell (41%)

Figure 2 shows the winning percentage of each team in 2006-2007 as a function of how frequently they yanked starters after threshold fouls in the third quarter. The better teams yanked more frequently.

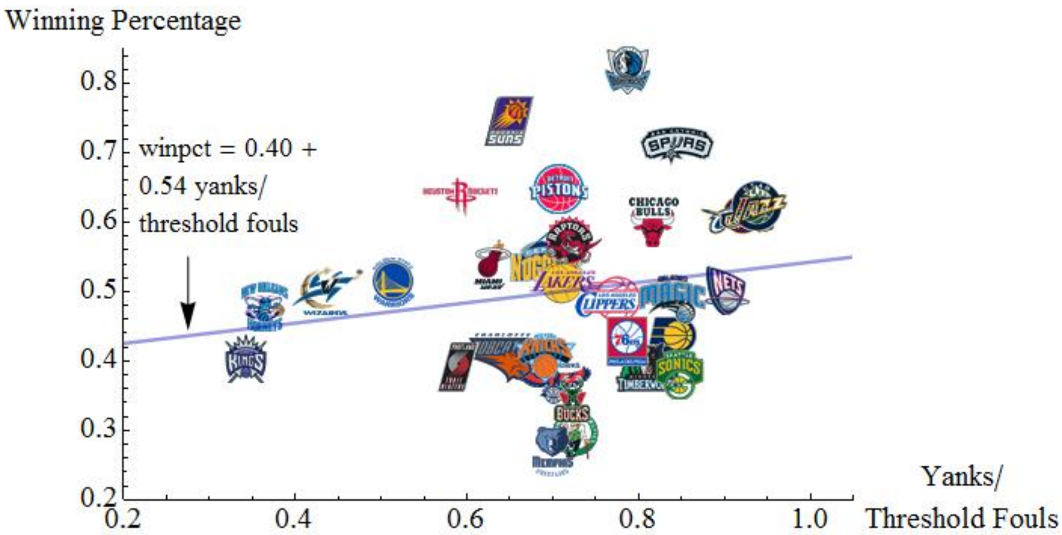


Figure 2: Winning Percentage in 2006-2007 versus Ratio of Yanks per Threshold Fouls in Q3

4.2. Win Probability Model

To give an indication of the dispersion in team coefficients, figure 3 displays the coefficients on the individual teams for 2006-2007, shown in bins. The Boston Celtics were the weakest team with a coefficient of -0.7 and the Dallas Mavericks were the strongest team with a coefficient of 1.1. Of course, the team coefficients for other years were different, particularly as the Boston Celtics had a league-leading coefficient of 0.9 in 2007-2008, but the general dispersion and ranges observed were similar.

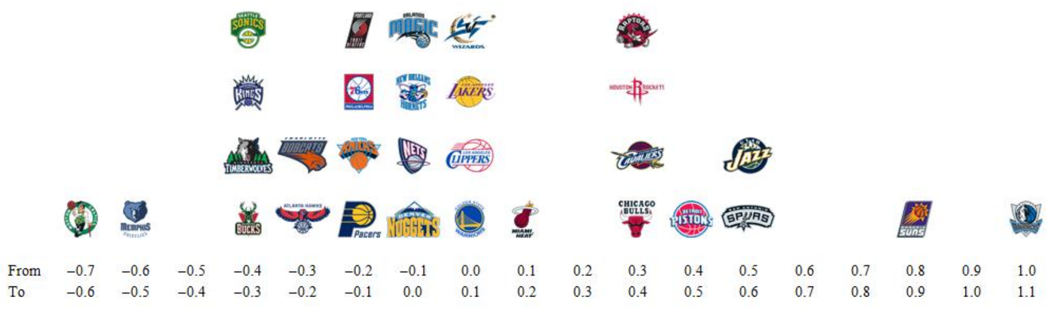


Figure 3: Team Coefficients for 2006-2007

Table 2 displays the rest of the regression coefficients for all three years individually and combined, along with the adjusted t-statistic in brackets below. The adjusted t-statistics are the raw t-statistics divided by 15.8 to account for the overlapping nature of the regressions. See Appendix A for more information.

Table 2: Coefficients and Adjusted t-Statistics of Win Probability Regression

	Constant α	Lead β_ℓ	Drift μ	Poss. β_P	Volatility γ	Starters β_{STA}	Fouls 1 β_{FTR}	Fouls 2 β_{cFTR}
2006-2007	-0.0195 [-0.91]	0.0543 [18.14]	0.2851 [4.95]	0.0278 [1.58]	0.9937 [1.21]	0.0176 [0.47]	-0.0619 [-0.29]	
2006-2007	-0.0194 [-0.91]	0.0543 [18.13]	0.2851 [4.94]	0.0279 [1.58]	0.9937 [1.21]	0.0174 [0.46]		-0.1860 [-0.38]
2007-2008	-0.0074 [-0.36]	0.0537 [18.21]	0.3461 [5.92]	0.0284 [1.61]	0.9980 [0.97]	-0.0194 [-0.50]	-0.1353 [-0.59]	
2007-2008	-0.0073 [-0.35]	0.0536 [18.19]	0.3464 [5.92]	0.0279 [1.58]	0.9979 [0.96]	-0.0192 [-0.49]		-0.3227 [-0.59]
2008-2009	0.0028 [0.13]	0.0548 [17.67]	0.3784 [6.32]	0.0301 [1.70]	0.9973 [0.81]	0.0141 [0.35]	-0.1295 [-0.55]	
2008-2009	0.0030 [0.14]	0.0547 [17.67]	0.3772 [6.30]	0.0299 [1.68]	0.9972 [0.81]	0.0140 [0.35]		-0.3100 [-0.60]
All 3 Years	-0.0093 [-0.76]	0.0553 [32.40]	0.3062 [9.23]	0.0290 [2.84]	0.9958 [1.78]	0.0104 [0.48]		-0.2983 [-1.03]

Note that the variance coefficient is below one, the assumption of Stern (1994).[3] The "time remaining" and "constant" coefficients combine with the team variables to tell us how a team's advantage evolves over time, with "constant" representing the end-of-game home court advantage. Surprisingly, in our data set the home court advantage disappears with three minutes remaining in the game.

The coefficient for foul trouble is negative each year, whether specified as FTR or cFTR.

5 Conclusions and Further Areas of Research

We describe a novel and general approach to estimating performance in professional basketball games by using tools from finance and we demonstrate its usefulness by applying it to the problem of early foul trouble, a problem that until now had been absent from the academic literature. Our analysis shows that most of the time, a starter in foul trouble should be yanked from the game. If left in the game, the player may become a liability, since he is afraid of picking up another foul. A possible area of research would be to examine how a player's statistics (points, rebounds, assists, steals, blocks, charges, plus/minus, etc.) are affected by foul trouble. Such analysis would complement our analysis. Also, we suggest that player fatigue and bench depth be incorporated into the analysis.

[3] The reported adjusted t-statistics for the coefficient of variance, γ, are reported testing the null hypothesis that $\gamma = 1$ rather than $\gamma = 0$ to match the expectation under a diffusion model without jumps.

Finally, the approach outlined here can likely be applied to numerous other questions such as comparing the efficacy of a three-guard lineup, testing for the effects of momentum or streaks, and even evaluating an individual's entire contribution, including defensively, by conditioning on their presence on the court.

6 Acknowledgments

The authors would like to thank Ashwin Alankar, David Berri, Jeffrey Chuang, Harry Gakidis, Cheng Ho, Alan Liu, Kenneth Shirley, Michael Shen, and Thomas Shen for their helpful comments.

References

[1] http://www.nba.com/2010/news/features/steve_aschburner/06/07/kobe.fouls/index.html

[2] http://sports.espn.go.com/nba/playoffs/2010/columns/story?columnist=adande_ja&page=Bryant-100607

[3] D. Rosenbaum, "Measuring How NBA Players Help Their Teams Win", http://www.82games.com/comm30.htm, 2004.

[4] H. Stern, "A Brownian Motion Model for the Progress of Sports Scores," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 1128 -1134, 1994.

[5] T. Moskowitz and L. Wertheim, Scorecasting: The Hidden Influences Behind How Sports are Played and Games are Won, 1st Edition, New York, NY, 2011.

[6] R. Merton, "Option Pricing When Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, vol. 3, pp. 125-144, 1976.

Appendix A

Stern [4] points out that approaches that estimate the probability of winning a game at various points throughout the game implicitly use overlapping data points. He reports simulation results that the t-statistics computed without adjusting for this overlapping phenomenon when the number of observations per game is $n = 4$ (i.e., observations at the end of each quarter) are overstated by 30 percent.

Figure A1 below displays the results of extending Stern's simulation approach to larger values of n. In the data set of this paper, using tick-by-tick data, the number of observations or ticks per game averages 498, so the appropriate t-statistic multiplier is 15.8.

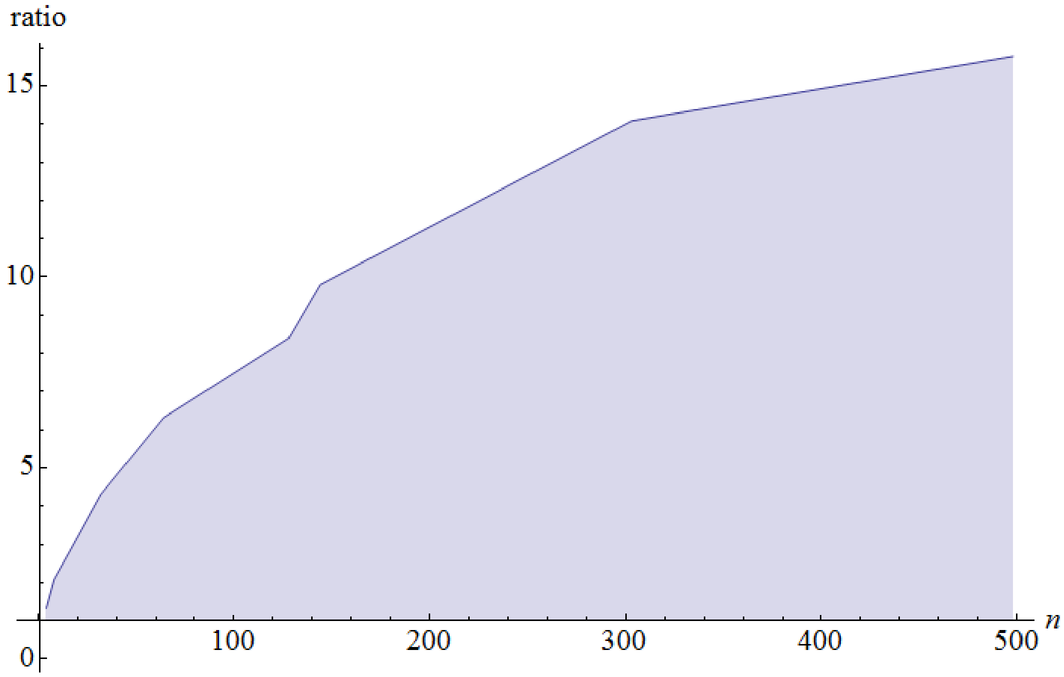


Figure A1: Stern Simulation t-statistic Adjustment Factor