Moneyballer: An Integer Optimization Framework for Fantasy Cricket League Selection and Substitution

Debarghya Das
dd367@cornell.edu

Abstract

This paper, Moneyballer [1] , uses binary integer programming to create optimal sequences of teams in fantasy sports leagues, particularly cricket. Some of the properties of team selection in sport are very similar to that in finance, which make this problem somewhat similar to the former. We recalibrate our model according to the distinctions between the two. For example, like the Black-Litterman model, our technique shows theoretical guarantees on overall team performance in a league dependent on your predicted return. Like the Markowitz model, it attempts to optimize on returns, however it does not penalize for risk and assumes (perhaps incorrectly), independence between performance of assets (players). Our model breaks software limits in previous research by **33x**, and produces a team which rates **99.53%ile** in a league of above 430,000 participants. It also introduces an automated backtesting framework for cricket performance measure testing, and supplies a method to fetch rich cricket statistics automatically.

1 Outlining the Problem

1.1 Cricket - Basic Gameplay

Cricket is a bat and ball game played by two teams of 11 players each. The game is played in several different variants. The format of the game we deal with, in particular, is the **T20 format**. Regarded as a fairly complicated game, here are some general gameplay rules of T20 cricket:

- One team goes out and bats, while the other bowls. This session is called an **innings**. When the first team's innings is finished, either because all their 11 players got **out** or because they finished playing their allotted time, the second team goes out to bat with the goal of getting as many runs as the first. Two batsmen bat together at a time. And if there is only one batsman left to bat, the innings is determined to be over.
- The time allotted per team in T20 cricket is **20 overs**. Each over consists of **6 balls** or deliveries. Each over is bowled by one member of the bowling team. Consecutive overs cannot be bowled by the same bowler. Typically, one innings takes one to one and a half hours.
- Each bowler can bowl **4 overs** in one innings. Typically 5 to 6 bowlers bowl from the bowling team in one innings to make up the entire 20 over quota.
- Two batsman from the batting team bat together at a time. They can hit the ball along the ground anywhere and run from one end to the other - a distance of 22 yards. For each time they run, they accumulate that many points.
- If the balls crosses the boundary line, a circle surrounding the batting area usually 60 - 80 meters away, without bouncing, the batsmen gets **6 points**, and if it does bounce, he gets **4 points**.
- There are several ways a batsman can get out.
  - He can be **caught** by any of the 11 people in the bowling side (including the bowler) without the ball touching the ground. When the bowler catches the ball of his own delivery, the dismissal is called **caught and bowled**.
  - He protects three sticks behind him when he bats. If he misses the ball and the ball hits any of the three sticks, he is out **bowled**.

[1] The title is a portmanteau of "moneyball" - a the book and following movie, *Moneyball*, which deals with the successful application of statistics in baseball to win 19 consecutive games with a below average financial budget `http://en.wikipedia.org/wiki/Moneyball`, and "baller" - a street ball player who has made it into the big leagues.

League Selection and Substitution

Debarghya Das dd367@cornell.edu

Abstract

This paper, Moneyballer 1 , uses binary integer program- ming to create optimal sequences of teams in fantasy sports leagues, particularly cricket. Some of the prop- erties of team selection in sport are very similar to that in finance, which make this problem somewhat similar to the former. We recalibrate our model according to the distinctions between the two. For example, like the Black-Litterman model, our technique shows theoretical guarantees on overall team performance in a league de- pendent on your predicted return. Like the Markowitz model, it attempts to optimize on returns, however it does not penalize for risk and assumes (perhaps incorrectly), independence between performance of assets (players). Our model breaks software limits in previous research by 33x, and produces a team which rates 99.53%ile in a league of above 430,000 participants. It also introduces an automated backtesting framework for cricket perfor- mance measure testing, and supplies a method to fetch rich cricket statistics automatically.

1 Outlining the Problem

1.1 Cricket - Basic Gameplay

Cricket is a bat and ball game played by two teams of 11 players each. The game is played in several differ- ent variants. The format of the game we deal with, in particular, is the T20 format. Regarded as a fairly com- plicated game, here are some general gameplay rules of T20 cricket:

• One team goes out and bats, while the other bowls. This session is called an innings. When the first

1The title is a portmanteau of ”moneyball” - a the book and fol- lowing movie, Moneyball, which deals with the successful applica- tion of statistics in baseball to win 19 consecutive games with a below average financial budget http://en.wikipedia.org/wiki/ Moneyball, and ”baller” - a street ball player who has made it into the big leagues.

team’s innings is finished, either because all their 11 players got out or because they finished playing their allotted time, the second team goes out to bat with the goal of getting as many runs as the first. Two batsmen bat together at a time. And if there is only one batsman left to bat, the innings is deter- mined to be over.

• The time allotted per team in T20 cricket is 20 overs. Each over consists of 6 balls or deliveries. Each over is bowled by one member of the bowl- ing team. Consecutive overs cannot be bowled by the same bowler. Typically, one innings takes one to one and a half hours.

• Each bowler can bowl 4 overs in one innings. Typ- ically 5 to 6 bowlers bowl from the bowling team in one innings to make up the entire 20 over quota.

• Two batsman from the batting team bat together at a time. They can hit the ball along the ground any- where and run from one end to the other - a distance of 22 yards. For each time they run, they accumu- late that many points.

• If the balls crosses the boundary line, a circle sur- rounding the batting area usually 60 - 80 meters away, without bouncing, the batsmen gets 6 points, and if it does bounce, he gets 4 points.

• There are several ways a batsman can get out.

– He can be caught by any of the 11 people in the bowling side (including the bowler) with- out the ball touching the ground. When the bowler catches the ball of his own delivery, the dismissal is called caught and bowled.

– He protects three sticks behind him when he bats. If he misses the ball and the ball hits any of the three sticks, he is out bowled.

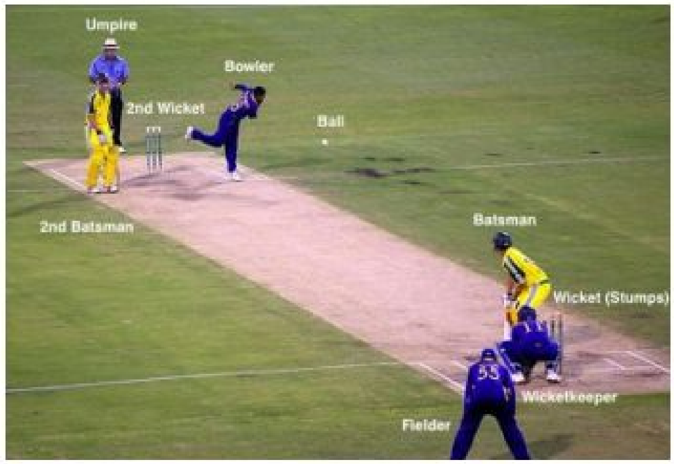

Figure 1: A snapshot of the typical action on a cricket field with some of the important characters labelled.

  - He also keeps some part of his foot behind a line called the crease near the sticks. If, while trying to play his shot, he moves out beyond that line, the person standing behind the sticks, called the wicketkeeper, can hit the sticks with the ball, to **stump** him. This is more uncommon than the previous two.
  - If he misses a ball which would have hit the stumps if he hadn't been there, then he gets dismissed **leg before wicket**.
  - While running from one end to another after hitting a ball, if any person from the opposing side throws the ball to the sticks at each end and hits it (directly or indirectly) before the batsman has crossed his line with some form of his body or his bat, then he is given out **run out**.

- If the first team puts up a score and either dismisses 10 players of the other team for a lower score, or the other team plays out their allotted overs without achieving their target, the first team wins. If the second team manages to score at least one more run than the first team, they are declared winners and the game ends.

1.2 Specific Rules and Restrictions

The specific Fantasy we use for testing is the biggest T20 tournament of the word, called the **Indian Premier League (IPL)**. Special rules apply to this tournament, and even more specifically to this Fantasy league, which are relevant to our problem:

- **Squad Balance** - Each player can have one of the following "skillets". There exist 3 "multiple skillet" combinations as well - Batsman + All Rounder, Batsman + Wicketkeeper, and Bowler + All Rounder:
  - **Batsman** (At least 4) A squad must have at least 4 specialist batsman. Although all 11 players can go bat, specialist batsmen typically score many more points, or runs, for his team.
  - **Bowler** (At least 2) A squad must have at least 2 specialist bowlers. Not all 11 players can bowl.
  - **All-Rounder** (At least 1) A squad must have at least 1 specialist all-rounder - some-one skilled at both batting and bowling.
  - **Wicketkeeper** (Exactly 1) A specialist role called the wicketkeeper has to be be in every team - at least one. He is the person who stands behind the sticks and collects the ball if the batsman misses, or attempts to catch balls that deflect off the bat, or stump the batsman when he steps out of his crease.
  - **Bowling Criteria** (At least 5) The sum of number of all rounders and bowlers on a particular team must be 5 or greater. Because each bowler can bowl 4 overs, and there are

Figure 1: A snapshot of the typical action on a cricket field with some of the important characters labelled.

– He also keeps some part of his foot behind a line called the crease near the sticks. If, while trying to play his shot, he moves out beyond that line, the person standing behind the sticks, called the wicketkeeper, can hit the sticks with the ball, to stump him. This is more uncom- mon than the previous two. – If he misses a ball which would have hit the stumps if he hadn't been there, then he gets dismissed leg before wicket. – While running from one end to another after hitting a ball, if any person from the opposing side throws the ball to the sticks at each end and hits it (directly or indirectly) before the batsman has crossed his line with some form of his body or his bat, then he is given out run out.

• If the first team puts up a score and either dismisses 10 players of the other team for a lower score, or the other team plays out their allotted overs with- out achieving their target, the first team wins. If the second team manages to score at least one more run than the first team, they are declared winners and the game ends.

1.2 Specific Rules and Restrictions

The specific Fantasy we use for testing is the biggest T20 tournament of the word, called the Indian Premier League (IPL). Special rules apply to this tournament,

and even more specifically to this Fantasy league, which are relevant to our problem:

• Squad Balance - Each player can have one of the following ”skillets”. There exist 3 ”multi- ple skillet” combinations as well - Batsman + All Rounder, Batsman + Wicketkeeper, and Bowler + All Rounder:

– Batsman (At least 4) A squad must have at least 4 specialist batsman. Although all 11 players can go bat, specialist batsmen typi- cally score many more points, or runs, for his team. – Bowler (At least 2) A squad must have at least 2 specialist bowlers. Not all 11 players can bowl. – All-Rounder (At least 1) A squad must have at least 1 specialist all-rounder - some-one skilled at both batting and bowling. – Wicketkeeper (Exactly 1) A specialist role called the wicketkeeper has to be be in ev- ery team - at least one. He is the person who stands behind the sticks and collects the ball if the batsman misses, or attempts to catch balls that deflect off the bat, or stump the batsman when he steps out of his crease. – Bowling Criteria (At least 5) The sum of number of all rounders and bowlers on a par- ticular team must be 5 or greater. Because each bowler can bowl 4 overs, and there are

20 overs in a match, every team needs at least 5 people who can bowl.

- **Overseas Limit** The Indian Premier League is primarily composed of domestic Indian players, with an abundance of foreign players from other top cricket playing nations, including England, Australia, South Africa and the West Indies.
  The format of the league restricts the number of overseas players in a playing squad to **4 players**.

- **Uncapped Quota** As one objective of the IPL is to foster domestic Indian players who have not played in any international matches before to rise to the occasion. In celebration of this, a squad must have **at least 1 uncapped player**

- **Franchise Spread** There are 8 teams in the IPL (at least this year). Each of them is based off of a particular location of the country. The 8 teams are:

  - **Kolkata Knight Riders (KKR)**
  - **Royal Challengers Bangalore (RCB)**
  - **Kings XI Punjab (KXIP)**
  - **Delhi Daredevils (DD)**
  - **Mumbai Indians (MI)**
  - **Rajasthan Royals (RR)**
  - **Sunrisers Hyderabad (SRH)**
  - **Chennai Super Kings (CSK)**

  A franchise spread criteria for Fantasy IPL is that you cannot have **more than 6 players** from one particular IPL franchise.

- **Financial Strength Criteria** Clearly, all the players available are not of equal strength - their skillets range from international legends, to upcoming uncapped 19 year olds. Based on their skill set, each player is given a price tag - ranging from **$600,000** to **$1,100,000**. The budget that you, as a manager, have while making your team

Figure 2: The logo of the 2014 Pepsi IPL T20

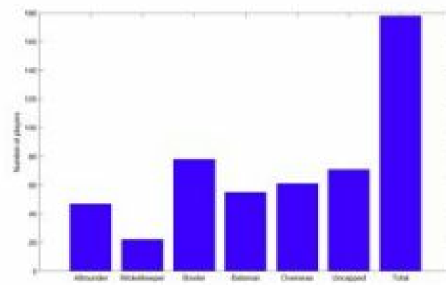

Figure 3: A distribution of the various roles of players in the IPL 2014.

- **Substitution Constraint** In the preliminary series of the tournament, there are **56 total games** played between the 8 teams. In each, a user receives points if a player he or she has is playing on that day gets a certain number of points depending on a scoring metric discussed later. If, for example, RCB plays KKR, and you have no RCB or KKR players in your team that day, you receive 0 points. Each user gets **75 substitutions** for the preliminary round.

- **Uncapped Player Substitution** To promote uncapped player, the Fantasy rules allow **one free uncapped player substitution** in every match, in addition to the 75.

- **Power Player** Every match, you get to choose your "power player". This player gets his points in the next match doubled.

- **Lock In** To prevent trading of players during the match, a user is only allowed to modify his team before the start of a match. After the match starts, his team is **locked**, and no further substitutions will count for points in that match.

1.3 Scoring System

The IPL 2014's preliminary stages consist of **56 games** from **April 16** to **May 25th**. On some days no games are played, but usually each day has **1-2** games. You are allowed infinite substitutions before the beginning of the tournament. Before the beginning of the next day's match, you're allowed to change your team. If a game X vs. Y occurs tomorrow at time t. At t, your team selection is locked down. After the game is finished, you receive points as per the rules above and system below. Points are only received by players on your fantasy that belong to teams X and Y if they played that day. All

20 overs in a match, every team needs at least 5 people who can bowl.

• Overseas Limit The Indian Premier League is pri- marily composed of domestic Indian players, with an abundance of foreign players from other top cricket playing nations, including England, Aus- tralia, South Africa and the West Indies. The format of the league restricts the number of overseas players in a playing squad to 4 players.

• Uncapped Quota As one objective of the IPL is to foster domestic Indian players who have not played in any international matches before to rise to the oc- casion. In celebration of this, a squad must have at least 1 uncapped player

• Franchise Spread There are 8 teams in the IPL (at least this year). Each of them is based off of a par- ticular location of the country. The 8 teams are:

– Kolkata Knight Riders (KKR) – Royal Challengers Bangalore (RCB) – Kings XI Punjab (KXIP) – Delhi Daredevils (DD) – Mumbai Indians (MI) – Rajasthan Royals (RR) – Sunrisers Hyderabad (SRH) – Chennai Super Kings (CSK)

A franchise spread criteria for Fantasy IPL is that you cannot have more than 6 players from one par- ticular IPL franchise.

• Financial Strength Criteria Clearly, all the play- ers available are not of equal strength - their skillets range from international legends, to upcoming un- capped 19 year olds. Based on their skill set, each player is given a price tag - ranging from $600,000 to $1,100,000. The budget that you, as a manager, have while making your team

Figure 2: The logo of the 2014 Pepsi IPL T20

Figure 3: A distribution of the various roles of players in the IPL 2014.

• Substitution Constraint In the preliminary series of the tournament, there are 56 total games played between the 8 teams. In each, a user receives points if a player he or she has is playing on that day gets a certain number of points depending on a scoring metric discussed later. If, for example, RCB plays KKR, and you have no RCB or KKR players in your team that day, you receive 0 points. Each user gets 75 substitutions for the preliminary round.

• Uncapped Player Substitution To promote un- capped player, the Fantasy rules allow one free un- capped player substitution in every match, in ad- dition to the 75.

• Power Player Every match, you get to choose your ”power player”. This player gets his points in the next match doubled.

• Lock In To prevent trading of players during the match, a user is only allowed to modify his team before the start of a match. After the match starts, his team is locked, and no further substitutions will count for points in that match.

1.3 Scoring System

The IPL 2014’s preliminary stages consist of 56 games from April 16 to May 25th. On some days no games are played, but usually each day has 1-2 games. You are allowed infinite substitutions before the beginning of the tournament. Before the beginning of the next day’s match, you’re allowed to change your team. If a game X vs. Y occurs tomorrow at time t. At t, your team selection is locked down. After the game is finished, you receive points as per the rules above and system below. Points are only received by players on your fantasy that belong to teams X and Y if they played that day. All

Figure 4: A look at the dashboard for exchange of players in the Fantasy IPL.

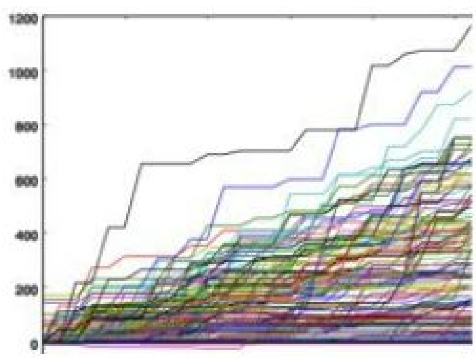

Figure 5: A graph showing the progression of points for each of the 178 players in the tournament.

players from other teams receive 0 points if they do not play on that day. After every day, you're allowed to make changes or substations to your side, but you are capped as per the rules above.

The fantasy league attempts to gauge the performance of distinct aspects of the game on one linear point scale. It uses various statistics based on a player's performance in a match to derive his points, which may be positive or negative. Note, the power player gets double his base points, regardless of whether negative or positive.

- **Batting points**
  - **Base Score** 1 point per run
    The more runs your score, the more points you get.
  - **Impact Score** 2 points per six, -5 per duck (getting out on 0 runs)
    In the smaller, 20 over format, six runs is a great boost in the scoring rate, and rewarded. Getting out without scoring is a disappointment, and penalized.
  - **Milestone Score** 10 points for every 25 runs scored
    Intuitively, 25, 50, 75 and 100 have been regarded as large milestones for runs in crickets in a given innings, and players are rewarded for it.
  - **Pace Bonus** Runs scored - balls faced
    Intuitively, players who hit more runs off fewer deliveries are more valuable, and they are rewarded thus.
- **Bowling points**
  - **Base Score** 20 points per wicket
    A wicket, or getting an opposing batsman out, is a huge prize. Only 10 can be taken in an innings, and each is rewarded with 20 points.

Figure 5: A graph showing the progression of points for each of the 178 players in the tournament.

players from other teams receive 0 points if they do not play on that day. After every day, you're allowed to make changes or substations to your side, but you are capped as per the rules above. The fantasy league attempts to gauge the performance of distinct aspects of the game on one linear point scale. It uses various statistics based on a player's performance in a match to derive his points, which may be positive or negative. Note, the power player gets double his base

Figure 4: A look at the dashboard for exchange of players in the Fantasy IPL.

points, regardless of whether negative or positive.

• Batting points

– Base Score 1 point per run

The more runs your score, the more points you get. – Impact Score 2 points per six, -5 per duck

(getting out on 0 runs) In the smaller, 20 over format, six runs is a great boost in the scoring rate, and rewarded. Getting out without scoring is a disappoint- ment, and penalized. – Milestone Score 10 points for every 25 runs

scored Intuitively, 25, 50, 75 and 100 have been re- garded as large milestones for runs in crickets in a given innings, and players are rewarded for it. – Pace Bonus Runs scored - balls faced

Intuitively, players who hit more runs off fewer deliveries are more valuable, and they are rewarded thus.

• Bowling points

– Base Score 20 points per wicket

A wicket, or getting an opposing batsman out, is a huge prize. Only 10 can be taken in an innings, and each is rewarded with 20 points.

- **Impact Score** 1 points per dot ball, 20 points for a maiden over
  A dot ball, or a ball where no run is scored, is rare and is a desirable by the bowler. A maiden over, or one segment of 6 consecutive dot balls, is rewarded handsomely with 20 points.
- **Milestone Score** 10 points for the 2nd wicket, 10 points for each subsequent wicket
  Taking more than one wicket in a T20 is an uncommon feat, and rewarded with bonuses.
- **Pace Bonus** 1.5x Balls bowled - runs conceded. If positive, it is doubled.
  A bowlers goal is to concede as few runs as possible, also known as maintaining a good economy rate. For doing this, they are rewarded with points, even if they fail to dismiss a batsman.

- **Fielding points**
  - **Catches** 10 points
    For a player who takes a catch to dismiss a batsman.
  - **Stumping** 15 points
    For a wicketkeeper who hits the sticks with the ball when the batsman is not within his crease, or his safe line.
  - **Direct Hit** 15 points
    When a fielder throws the ball from the distance and it directly hits the sticks while the batsmen are running, and they are not in their safe zone, then they are rewarded points.
  - **Run Out** 10 points for each player involved
    Similar to a direct hit, except the throw from the distance need not directly hit the stumps. As long as a batsman is dismissed before he makes his ground, all fielding players involved in throwing the ball are rewarded.
- **Bonus points**
  - **Player of the match** 25 points
    The most valuable player of every match is rewarded 25 more points for his performance.
  - **Victory Bonus** 5 points
    A player in the victorious side is rewarded with 5 points.

1.4 The Problem

We broadly separate our problem in to 2 parts - **static prediction**, which creates a selection given rules and constraints on day 1 for the entire fantasy league without updating expectations after the actual cricket games take place, and **dynamic prediction**, which solves the problem after every day of play, updating its expected player performance with time.

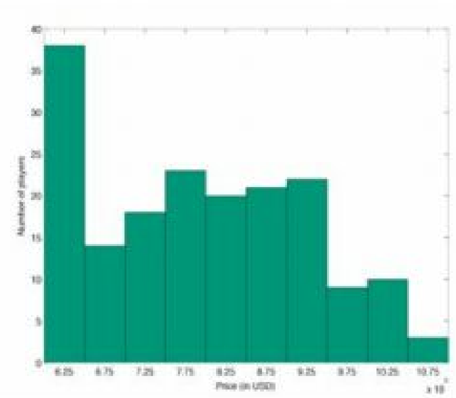

Figure 7: A distribution of the prices of all the players in the IPL. They range from $600k to $1.1m.

1.4.1 Static Prediction

In plain words, our problem S is to pick a team of 11 players for every one of the 37 days of play, each of which should abide by certain cricketing squad balance and budgeting based team constraints. The total number of substitutions incurred in the selection of all these teams should be less than equal to 75. The resultant selection of teams should maximize the expected points that you, as a "manager" collect from the entire endeavor. The point formula is strongly correlated with the general quality of cricket performances. The initial expectations of all the players can be given by e, in which case this problem problem produces a solution $S(e)$. The static forecast is a solvable problem, with a "correct" solution to be found.

1.4.2 Dynamic Prediction

If the Static Prediction problem is called S, then the Dynamic Prediction problem solves S after each day of play with an expectation e. If e were to remain constant from day to day, then the solution of $S_1(e)$, the static prediction given an expectation e on day 1, would be such that $S_2(e) \in S_1(e)$. If it remained constant for all n days,

$$S_{k+1}(e) \in S_k(e) \;\; \forall 1 \leq k < n \;\; , k \in \mathbb{Z}$$

– Impact Score 1 points per dot ball, 20 points

for a maiden over A dot ball, or a ball where no run is scored, is rare and is a desirable by the bowler. A maiden over, or one segment of 6 consecu- tive dot balls, is rewarded handsomely with 20 points. – Milestone Score 10 points for the 2nd wicket,

10 points for each subsequent wicket Taking more than one wicket in a T20 is an uncommon feat, and rewarded with bonuses. – Pace Bonus 1.5x Balls bowled - runs con-

ceded. If positive, it is doubled. A bowlers goal is to concede as few runs as possible, also known as maintaining a good economy rate. For doing this, they are re- warded with points, even if they fail to dismiss a batsman.

• Fielding points

– Catches 10 points

For a player who takes a catch to dismiss a batsman. – Stumping 15 points

For a wicketkeeper who hits the sticks with the ball when the batsman is not within his crease, or his safe line. – Direct Hit 15 points

When a fielder throws the ball from the dis- tance and it directly hits the sticks while the batsmen are running, and they are not in their safe zone, then they are rewarded points. – Run Out 10 points for each player involved

Similar to a direct hit, except the throw from the distance need not directly hit the stumps. As long as a batsman is dismissed before he makes his ground, all fielding players involved in throwing the ball are rewarded.

• Bonus points

– Player of the match 25 points

The most valuable player of every match is re- warded 25 more points for his performance. – Victory Bonus 5 points

A player in the victorious side is rewarded with 5 points.

1.4 The Problem

We broadly separate our problem in to 2 parts - static prediction, which creates a selection given rules and

Figure 7: A distribution of the prices of all the players in the IPL. They range from $600k to $1.1m.

constraints on day 1 for the entire fantasy league without updating expectations after the actual cricket games take place, and dynamic prediction, which solves the prob- lem after every day of play, updating its expected player performance with time.

1.4.1 Static Prediction

In plain words, our problem S is to pick a team of 11 players for every one of the 37 days of play, each of which should abide by certain cricketing squad balance and budgeting based team constraints. The total num- ber of substitutions incurred in the selection of all these teams should be less than equal to 75. The resultant se- lection of teams should maximize the expected points that you, as a ”manager” collect from the entire endeavor. The point formula is strongly correlated with the general quality of cricket performances. The initial expectations of all the players can be given by e, in which case this problem problem produces a solution S(e). The static forecast is a solvable problem, with a ”correct” solution to be found.

1.4.2 Dynamic Prediction

If the Static Prediction problem is called S, then the Dy- namic Prediction problem solves S after each day of play with an expectation e. If e were to remain constant from day to day, then the solution of $S_1(e)$, the static predic- tion given an expectation e on day 1, would be such that $S_2(e) \in S_1(e)$. If it remained constant for all n days,

$$S_{k+1}(e) \in S_k(e) \; \forall 1 \le k < n, k \in \mathbb{Z}$$

Figure 6: The 8 teams participating in the IPL 2014.

Therefore, applying each static forecast solution on each day would be equivalent to applying $S_1(e)$.
However, in sport, the odds of all players performing exactly as you expect is negligible. As such, we have a separate e_k for every game. If $S_k(e)$ produces a list of teams for days $k...n$, we refer to the i^{th} team that $S_k(e)$ predicts as $S_k(e)[i]$ The Dynamic Prediction algorithm D is simply

$$D = [S_k(e_k))[k] \mid 1 \leq k \leq n]$$

This is simply an aggregate of static predictions for updated player performance expectations.

2 Why this is an interesting problem?

Intellectual Merit

- **Uniqueness** The area of using optimization based approaches for team selection in sport, especially cricket, has not been a typically wide area of research. The only previous research on it varies little, and comes from a Econometrics and Sports Modeling research group at *Nelson Mandela Metropolitan University* [2], with Gerber and Sharp [2006] being the seminal work in the field.

- **Directly applicable to Operations Research** While typically most of the work in this field is limited to Sports journals, or Statistical journals, this particular approach to the problem has been published in *Journal of the Operational Research Society* [Sharp et al., 2011].

- **The use of an NP-complete 0-1 Integer Linear Programming Method** In computational complexity theory, Karp [1972] reduced 21 combinatorial problems, including 0-1 linear integer programming, to SAT, proving it NP-complete. This makes the problem a *superpolynomial* in the input size, and a challenging problem to solve.

- **Using *Gurobi* to its limits** In Meindl and Templ [2012], we learn that Gurobi ranks 1st in all standard benchmarks [Dolan and Moré, 2002], Mittelmann LP benchmarks, Mittelmann [2003]and 2nd amongst all other benchmarks set by the paper[3] in performance while solving linear problems. Gurobi [4] is a commercial optimization software used to solve a wide variety of optimization problems. Most noticeably, Gurobi has multi-core parallelism features and a concise API. One article says "The harder the problem, they say, the better their performance and the bigger the gap." [5].
The cutting edge performance of Gurobi makes it an ideal solution to a problem which was once previously bounded by software limits.

- **Similarity and applicability to the classic financial trading models** The general framework of the problem fits in with the algorithms traditionally used in finance - such as the Markowitz Model [Markowitz, 1952] and the Model [He, 1999]. We discuss in further sections.

- **Making available fairly inaccessible data** Previous approaches [Brettenny, 2010] use manual data collection methods using *Excel*. We automate data collection and updating using a headless browser like PhantomJS and certain jQuery selectors to scrape recent data [6].
Our superior data collection method not only lends itself to a more comprehensive, automated framework, but provides a cohesive set of data for backtesting as the tournament progresses. This backtesting data could be invaluable for future benchmark-

[2] GD Sharp, WJ Brettenny, JW Gonsalves, M Lourens & RJ Stretch

[3] The software that ranked first was CPLEX, developed by two of the three Gurobi founders.

[4] `http://www.gurobi.com/`

[5] "First Look - Gurobi Optimization" - James Taylor - `http://jtonedm.com/2011/03/02/first-look-gurobi-optimization/`

[6] We currently have the Javascript for jQuery selectors, but have not integrated the headless browser into the update framework

Figure 6: The 8 teams participating in the IPL 2014.

Therefore, applying each static forecast solution on each day would be equivalent to applying S

1

• Using Gurobi to its limits In Meindl and Templ (e).

[2012], we learn that Gurobi ranks 1st in all

stan- However, in sport, the odds of all players performing ex-

dard benchmarks [Dolan and Mor e, 2002],

Mittel- actly as you expect is negligible. As such, we have a

mann LP benchmarks, Mittelmann [2003]and

2nd separate e

k for every teams for days k...n, we game. refer If to S

k the (e) ith produces team that a list of

S

k

(e) predicts as S

k

(e)[i] The Dynamic Prediction algorithm D

amongst all other benchmarks set by the paper3 in performance while solving linear problems. Gurobi 4 is a commercial optimization software used to is simply

solve a wide variety of optimization problems. Most D = [S

k

(e

k

))[k] | 1 ▯ k ▯ n]

This is simply an aggregate of static predictions for up- dated player performance expectations.

noticeably, Gurobi has multi-core parallelism fea- tures and a concise API. One article says ”The harder the problem, they say, the better their per- formance and the bigger the gap.” 5. The cutting edge performance of Gurobi makes it 2 Why this is an interesting problem?

an ideal solution to a problem which was once pre- viously bounded by software limits. Intellectual Merit

• Similarity and applicability to the classic finan-

• Uniqueness The area of using optimization based

cial trading models The general framework of the

approaches for team selection in sport, especially

problem fits in with the algorithms traditionally

cricket, has not been a typically wide area of re-

used in finance - such as the Markowitz Model

search. The only previous research on it varies little,

[Markowitz, 1952] and the Model [He, 1999]. We and comes from a Econometrics and Sports Model-

discuss in further sections. ing research group at
Nelson Mandela Metropolitan University 2, with Gerber and Sharp [2006] being

• Making available fairly inaccessible data Pre- the
seminal work in the field.

vious approaches [Brettenny, 2010] use manual data collection methods using Excel. We auto-

• Directly applicable to Operations Research

mate data collection and updating using a headless
While typically most of the work in this field is lim-

browser like PhantomJS and certain jQuery selec-
ited to Sports journals, or Statistical journals, this

tors to scrape recent data 6. particular approach to
the problem has been pub-

Our superior data collection method not only lends
lished in Journal of the Operational Research Soci-

itself to a more comprehensive, automated frame-
ety [Sharp et al., 2011].

work, but provides a cohesive set of data for back-

• The use of an NP-complete 0-1 Integer Linear Programming Method In computational complex-

testing as the tournament progresses. This backtest- ing data could be invaluable for future benchmark-

ity theory, Karp [1972] reduced 21 combinato-

3The software that ranked first was CPLEX,
developed by two of rial problems, including 0-1 linear integer program- ming, to SAT, proving it
NP-complete. This makes the problem a superpolynomial in the input size, and a challenging problem to solve.

the three Gurobi founders.

4http://www.gurobi.com/ 5"First Look - Gurobi Optimization" - James Taylor - http://jtonedm.com/2011/03/02/first-look-gurobi-optimization/

6We currently have the Javascript for jQuery selectors,
but have not 2GD Sharp, WJ Brettenny, JW Gonsalves, M Lourens & RJ Stretch

integrated the headless browser into the update framework ing of predictive performance metrics, which is the majority of cricket research, as discussed in the next section.

Broader Impact

- **Predictive Application to Fantasy Sports Selection** Fantasy sports leagues has become extremely popular. The English Premier League Fantasy Sports League has over 2,000,000 participants [Liantzas et al., 2013] and the IPL Fantasy League, in 2014, has around 432,000 participants. [7]
  This paper provides a cohesive framework that can be used an average Fantasy League player, much like an investor in finance, using either performance measures supplied by the algorithm, or their own instinct, such as in Black-Litterman. Regardless of what they choose, the framework guarantees an optimal set of teams and substitution, leaving the burden of satisfying constraints on the software. The user merely has to write an algorithm or use instinct to input his predictions on player performance, or use one pre-written our system.

- **Generalizable Application to Constrained Combinatorial Selection** Our framework does a lot of things out of the box - automatic up-to-date data retrieval, backtesting, and solves large optimization problems (using Gurobi) to select teams. The only things to modify are - schedule to backtest on, data for players, rules, and the endpoint from which to update data (or a historical store). It can be used for many more applications than merely cricket.

- **Direct Application to Statistical Cricket Selection** Selection into sports teams is a matter of pride and competition. Very often, we hear of controversial sports selections, that cause an uproar. We want to back the intuition behind team formation and selection with a theoretical, statistical framework.
  Although, the particular constraints restrict the software to the IPL Fantasy League, it can trivially be modified to asset in statistical based selection for real world international, or even by teams in the IPL, to assist in which players to buy in auctions for how much and how they will perform. With some modifications to the framework, application to a real-world cricket scenario is not difficult.

3 Previous Works

There exists a fairly small amount of total papers based on computational cricket in statistics, many of which are discussed below. These works can be broadly categorized into several groups, and I discuss how their contributions are relevant yet fall short of solving the problem at hand:

- **Data Aggregation and Performance Indicators** An overwhelming amount of work in the sports statistics area, particularly cricket, is done purely on aggregating and visualizing statistics. In Damodaran [2006], the author aggregates and visualizes certain data metrics of around 14 players, without making any key insights. Most papers are centered solely on finding statistics that are most indicative of success of a particular type of player in a particular sport.
  Saikia et al. [2012] uses a fairly complex machine learning method using Artificial Neural Networks for predictive performance of bowlers.
  Sharma et al. [2012] attempts to pick the best batsmen in T20 with an ordered weighted average while Amin and Sharma [2014] tries a more in-depth approach two-way regression approach.
  Lemmer [2011] attempts to find statistically significant performance measures for wicketkeepers in cricket.
  Sharma [2013] does a general study of performance indicators in the sport.
  These works are all very useful for the future, and could all be integrated into my model as a performance predictor.

- **Singular Team Selection with Little Predictive Ability** There exists a large amount of work that assists in selection of a single team in cricket. They aren't meant to be used in the same setting as the algorithm in this paper. They are not-self updating and are not built into backtesting frameworks.
  For example, in Omkar and Verma [2003], a genetic algorithm was used to create a cricket team. The study seemed to optimize one team of 11 players from a set of 23, based on certain "fitness" parameter. It does not indicate sufficiently adequately results regarding team balance. Neither is it sufficiently adaptive from one game to the next. It does not discuss how useful this "fitness" parameter is in predicting future performance. In Ahmed et al. [2011], they use a multi-objective optimization function to optimization on bowling averages and batting averages of players. This approach has many flaws. For example, it assumes that batting and bowling averages are the best metrics to optimize on to construct the perfect team. Secondly, the manner in which the paper tries to construct appropriate teams from the "knee cap" of the Pareto-optimal front seems like a add-on sanity check. Fur-

[7] Accessed 17th May, 2014. `https://fantasy.iplt20.com/ifl/leaderboard/list/allteams`

ing of predictive performance metrics, which is the majority of cricket research, as discussed in the next

section.

Broader Impact

• Predictive Application to Fantasy Sports Se- lection Fantasy sports leagues has become ex- tremely popular. The English Premier League Fan- tasy Sports League has over 2,000,000 participants [Liantzas et al., 2013] and the IPL Fantasy League, in 2014, has around 432,000 participants. 7 This paper provides a cohesive framework that can be used an average Fantasy League player, much like an investor in finance, using either performance measures supplied by the algorithm, or their own instinct, such as in Black-Litterman. Regardless of what they choose, the framework guarantees an op- timal set of teams and substitution, leaving the bur- den of satisfying constraints on the software. The user merely has to write an algorithm or use instinct to input his predictions on player performance, or use one pre-written our system.

• Generalizable Application to Constrained Com- binatorial Selection Our framework does a lot of things out of the box - automatic up-to-date data re- trieval, backtesting, and solves large optimization problems (using Gurobi) to select teams. The only things to modify are - schedule to backtest on, data for players, rules, and the endpoint from which to update data (or a historical store). It can be used for many more applications than merely cricket.

• Direct Application to Statistical Cricket Selec- tion Selection into sports teams is a matter of pride and competition. Very often, we hear of controver- sial sports selections, that cause an uproar. We want to back the intuition behind team formation and se- lection with a theoretical, statistical framework. Although, the particular constraints restrict the soft- ware to the IPL Fantasy League, it can trivially be modified to asset in statistical based selection for real world international, or even by teams in the IPL, to assist in which players to buy in auctions for how much and how they will perform. With some modifications to the framework, application to a real-world cricket scenario is not difficult.

3 Previous Works

There exists a fairly small amount of total papers based on computational cricket in statistics, many of which are

7Accessed 17th May, 2014. https://fantasy.iplt20.com/ ifl/leaderboard/list/allteams

discussed below. These works can be broadly catego- rized into several groups, and I discuss how their contri- butions are relevant yet fall short of solving the problem at hand:

• Data Aggregation and Performance Indicators An overwhelming amount of work in the sports statistics area, particularly cricket, is done purely on aggregating and visualizing statistics. In Damodaran [2006], the author aggregates and vi- sualizes certain data metrics of around 14 players, without making any key insights. Most papers are centered solely on finding statistics that are most in- dicative of success of a particular type of player in a particular sport. Saikia et al. [2012] uses a fairly complex machine learning method using Artificial Neural Networks for predictive performance of bowlers. Sharma et al. [2012] attempts to pick the best bats- men in T20 with an ordered weighted average while Amin and Sharma [2014] tries a more in-depth ap- proach two-way regression approach. Lemmer [2011] attempts to find statistically signif- icant performance measures for wicketkeepers in cricket. Sharma [2013] does a general study of performance indicators in the sport. These works are all very useful for the future, and could all be integrated into my model as a perfor- mance predictor.

• Singular Team Selection with Little Predictive Ability There exists a large amount of work that as- sists in selection of a single team in cricket. They aren’t meant to be used in the same setting as the algorithm in this paper. They are not-self updating and are not built into backtesting frameworks. For example, in Omkar and Verma [2003], a ge- netic algorithm was used to create a cricket team. The study seemed to optimize one team of 11 play- ers from a set of 23, based on certain ”fitness” pa- rameter. It does not indicate sufficiently adequately results regarding team balance. Neither is it suf- ficiently adaptive from one game to the next. It does not discuss how useful this ”fitness” parame- ter is in predicting future performance. In Ahmed et al. [2011], they use a multi-objective optimiza- tion function to optimization on bowling averages and batting averages of players. This approach has many flaws. For example, it assumes that batting and bowling averages are the best metrics to opti- mize on to construct the perfect team. Secondly, the manner in which the paper tries to construct ap- propriate teams from the ”knee cap” of the Pareto- optimal front seems like a add-on sanity check. Fur-

ther, it only models one team of 11 players , and chooses from a pool of 129 players.
Similarly, works like Bhattacharjee and Saikia [2013] and Lemmer [2013] purely deal with rather arbitrary measures for the far more simple task of simple selection.
Summers et al. [2007] was a pioneering work in statistical cricket selection, used in hockey leagues.

- **0-1 Integer Optimization for Modelling Team Selection and Substitution** Of all the areas of computational cricket research, this area strikes me as most relevant and closest to the work that I present [8].
As stated before, there seems to be only one Econometrics and Sports Modeling research group that has done all the work in this area - Brettenny [2010], Lourens [2009], Brettenny et al. [2012], Sharp et al. [2011] - all talk about the same concept, similar to my approach. The seminal work in this subfield is Gerber and Sharp [2006], also from this research group. However, all these papers have key setbacks -
  - They have a manual and limited data collection process. They capture data from only one previous tournament, and only compute aggregate, and not match by math statistics.
  - Their software limits them to operating on 50 players for every matches for 4 matches total.
  - They do not develop a framework for backtesting the model.
  - They cannot be practically used in a Fantasy League environment.
  - They conform to the older Fantasy League standards, which are riddled with several overtly lax, or unreasonable constraints on cricket gameplay. The Indian Premier League started in 2008, and most of their research conforms to standards between 2008-2010. This paper uses the updated rules and scoring methods of the 7th season, [9]. Several of these newer standards are trickier to integrate into the integer optimization framework.

4 Relationship to Finance

In many ways, fantasy league player selection and expected point optimization is similar to well studied finance theories.

[8]At the time of writing this paper, I had no clue of the existence of these works. It turns out that I solve a very similar problem in a similar way, but in ways discussed throughout the paper, better.

[9]Official IPL T20 Fantasy League Standards, 2014 - `https://fantasy.iplt20.com/ifl/default/faq#`

4.1 Similarities

- **We have a portfolio we want to optimize** In the finance analogy, assets are players, and portfolios are teams we select at different time stamps. We want to optimize the performance of this changing portfolio over time (this becomes our objective function).
- **We can model an expected return based on past performance** Similar to the Markowitz model [Markowitz, 1952], we can model performance of our assets based on past performance. Also, similarly, these models are usually not consistently accurate.
- **We have a budget restriction on the assets we purchase** In finance, we have a specific amount of wealth we can choose to invest in our portfolio. Similarly, in the fantasy league, we have a specific budget restriction for selecting our portfolio.
- **Integrating (better) Investor beliefs is possible, and sought** In the Black-Litterman model [He, 1999], investor beliefs are used to guide the expected value of the assets. Similarly, in fantasy cricket, because the past predictors are fairly week, expert investor beliefs can trivially be integrated to model expected return.

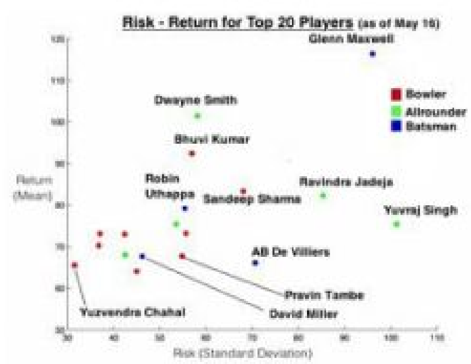

Figure 8: The demonstration of risk-return tradeoffs between 20 top IPL 2014 performers, according to their role in the team, or analogously, their "asset class".

- **There exist a risk-return tradeoff between different asset classes** In finance, for example, you have 2 distinct types of asset classes - bonds and stocks. Bonds have low risk and low return, while stocks have high risk and higher return. In cricket, player roles work in similar ways. Because bowlers always

ther, it only models one team of 11 players , and chooses from a pool of 129 players. Similarly, works like

Bhattacharjee and Saikia [2013] and Lemmer [2013] purely deal with rather arbitrary measures for the far more simple task of simple selection. Summers et al. [2007] was a pioneering work in sta- tistical cricket selection, used in hockey leagues.

• 0-1 Integer Optimization for Modelling Team Se- lection and Substitution Of all the areas of compu- tational cricket research, this area strikes me as most relevant and closest to the work that I present 8. As stated before, there seems to be only one Econo- metrics and Sports Modeling research group that has done all the work in this area - Brettenny [2010], Lourens [2009], Brettenny et al. [2012], Sharp et al. [2011] - all talk about the same concept, similar to my approach. The seminal work in this subfield is Gerber and Sharp [2006], also from this research group. However, all these papers have key setbacks -

– They have a manual and limited data collec- tion process. They capture data from only one previous tournament, and only compute ag- gregate, and not match by math statistics. – Their software limits them to operating on 50 players for every matches for 4 matches total. – They do not develop a framework for backtest-

ing the model. – They cannot be practically used in a Fantasy

League environment. – They conform to the older Fantasy League standards, which are riddled with several overtly lax, or unreasonable constraints on cricket gameplay. The Indian Premier League started in 2008, and most of their research conforms to standards between 2008-2010. This paper uses the updated rules and scoring methods of the 7th season, 9. Several of these newer standards are trickier to integrate into the integer optimization framework.

4 Relationship to Finance

In many ways, fantasy league player selection and ex- pected point optimization is similar to well studied fi- nance theories.

8At the time of writing this paper, I had no clue of the existence of these works. It turns out that I solve a very similar problem in a similar way, but in ways discussed throughout the paper, better.

9Official IPL T20 Fantasy League Standards, 2014 - https:// fantasy.iplt20.com/ifl/default/faq#

Figure 8: The demonstration of risk-return tradeoffs be- tween 20 top IPL 2014 performers, according to their role in the team, or analogously, their ”asset class”.

• There exist a risk-return tradeoff between differ- ent asset classes In finance, for example, you have 2 distinct types of asset classes - bonds and stocks. Bonds have low risk and low return, while stocks have high risk and higher return. In cricket, player roles work in similar ways. Because bowlers always

4.1 Similarities

• We have a portfolio we want to optimize In the fi- nance analogy, assets are players, and portfolios are teams we select at different time stamps. We want to optimize the performance of this changing portfolio over time (this becomes our objective function).

• We can model an expected return based on past performance Similar to the Markowitz model [Markowitz, 1952], we can model performance of our assets based on past performance. Also, simi- larly, these models are usually not consistently ac- curate.

• We have a budget restriction on the assets we purchase In finance, we have a specific amount of wealth we can choose to invest in our portfolio. Similarly, in the fantasy league, we have a specific budget restriction for selecting our portfolio.

• Integrating (better) Investor beliefs is possible, and sought In the Black-Litterman model [He, 1999], investor beliefs are used to guide the ex- pected value of the assets. Similarly, in fantasy cricket, because the past predictors are fairly week, expert investor beliefs can trivially be integrated to model expected return.

bowl regardless of game events, and great bowlers always return a sizable number of points, they are like bonds. The way we model bowler returns is very different from batsmen. Even great batsmen can get out on a duck (without scoring runs), or alternatively score many runs very quickly, and acquire much more points than bowlers will ever be able to. They are like stocks.

- **We can potentially model riskiness in assets** Similar to finance, it is not trivial to accurately describe the riskiness of an asset. Yet with similar standard deviation metrics, it is possible to describe an asset's riskiness, and this can further be optimized in the objective function.

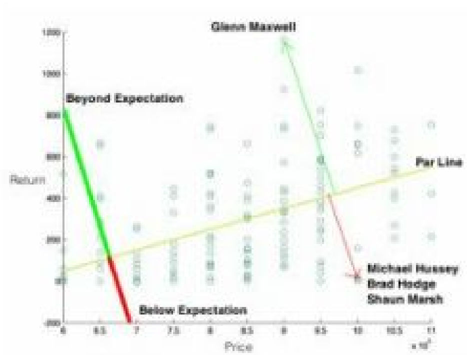

Figure 9: The demonstration of how player performances have stacked up to their price, showing the most over-priced and under-priced players of the tournament.

4.2 Differences

- **We cannot short assets** In finance, we can bet against an asset by short selling it. Unfortunately, in fantasy sports league, you can only buy assets (players), not bet against them.

- **Changes to the portfolio aren't charged, they are restricted** Unlike in finance, where typically changes to the portfolio are priced with a transaction cost, either fixed or variable or both. In fantasy cricket, the total number of changes you can make to your portfolio are restricted, not charged.

- **More than one "amount" of an asset can't be held** In finance, we can hold any amount of a certain asset. In fantasy sport, we cannot own a player

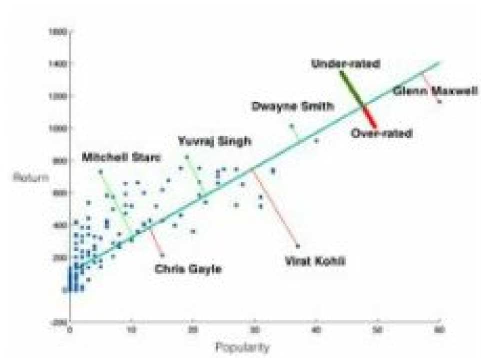

Figure 10: The demonstration of how player performances have stacked up to their popularity amongst IPL fans and fantasy players, showing the most over-rated and under-rated players of the tournament.

more than once [10].

- **Assets are assumed to be negligibly correlated** Whereas in finance, you have sets of assets who's prices are strongly correlated, and the market tends to trend in one direction, in fantasy cricket, one assumption (debatable) is that the assets move independently of one another.

- **All Assets do not change price everyday, but we know which ones do when** Whereas in finance, every asset moves in value every day, in fantasy cricket, the value of a player changes only if he plays a match. Otherwise, their scores are guaranteed to remain constant.

- **Tomorrow's asset price expectation is not centered on todays** In finance, we express the gain or loss as a function of yesterday's asset price. Everything moves in percentages about the center.
  In cricket, the batsman or bowler starts from scratch in every match. While there is a slight correlation between a good performance in the match before and this match, a consistency element called "form", a player scores do not tend to their score in the last match.

- **The returns on our assets are a distinct value from the spending on them** In finance, when you invest in assets and make a profit, the profit can be

[10] A point to note is that the IPL T20 Fantasy Cricket League does have a facility for setting a **power player** on every game. This power player gets rewarded with double his or her original score. In some sense, 2 of the same asset is held.

Figure 9: The demonstration of how player performances have stacked up to their price, showing the most

over- priced and under-priced players of the tournament.

4.2 Differences

• We cannot short assets In finance, we can bet against an asset by short selling it. Unfortunately, in fantasy sports league, you can only buy assets (players), not bet against them.

• Changes to the portfolio aren't charged, they are restricted Unlike in finance, where typically changes to the portfolio are priced with a transac- tion cost, either fixed or variable or both. In fantasy cricket, the total number of changes you can make to your portfolio are restricted, not charged.

• More than one "amount" of an asset can't be held In finance, we can hold any amount of a cer- tain asset. In fantasy sport, we cannot own a player

bowl regardless of game events, and great bowlers always return a sizable number of points, they are like bonds. The way we model bowler returns is very different from batsmen. Even great batsmen can get out on a duck (without scoring runs), or al- ternatively score many runs very quickly, and ac- quire much more points than bowlers will ever be able to. They are like stocks.

• We can potentially model riskiness in assets Sim- ilar to finance, it is not trivial to accurately describe the riskiness of an asset. Yet with similar standard deviation metrics, it is possible to describe an as- set's riskiness, and this can further be optimized in the objective function.

Figure 10: The demonstration of how player perfor- mances have stacked up to their popularity amongst IPL fans and fantasy players, showing the most over-rated and under-rated players of the tournament.

more than once 10.

• Assets are assumed to be negligibly correlated Whereas in finance, you have sets of assets who's prices are strongly correlated, and the market tends to trend in one direction, in fantasy cricket, one as- sumption (debatable) is that the assets move inde- pendently of one another.

• All Assets do not change price everyday, but we know which ones do when Whereas in finance, every asset moves in value every day, in fantasy cricket, the value of a player changes only if he plays a match. Otherwise, their scores are guaran- teed to remain constant.

• Tomorrow's asset price expectation is not cen- tered on todays In finance, we express the gain or loss as a function of yesterday's asset price. Every- thing moves in percentages about the center. In cricket, the batsman or bowler starts from scratch in every match. While there is a slight correla- tion between a good performance in the match be- fore and this match, a consistency element called "form", a player scores do not tend to their score in the last match.

• The returns on our assets are a distinct value from the spending on them In finance, when you invest in assets and make a profit, the profit can be

10A point to note is that the IPL T20 Fantasy Cricket League does have a facility for setting a power player on every game. This power player gets rewarded with double his or her original score. In some sense, 2 of the same asset is held.

reinvested into your portfolio. In fantasy cricket, your reward and spending are separate entities. No matter how much your reward, you will always be restricted to the same budget restrictions as before throughout the tournament.

- **Prices of assets do not change** Continuing from last point, the price of the asset also doesn't change in fantasy cricket.
- **Negative returns are rare** Although small negative returns are theoretically possible, they are few and far between.

These key difference force us to adapt our finance model to fit the criteria.

5 Math & Methodology

5.1 Theory

We attempt to formulate our problem as a mathematical optimization problem. Most mathematical optimization problems have 3 key components:

- **Decision Variables** These variables make the ultimate decision of the optimization problem.
- **Constraints** These are the restrictions imposed upon the decision variables.
- **Objective Function** It defines what must be optimized, in terms of the decision variables.

These 3 components are essential to formulate an optimization problem. Particularly relevant to our approach is Linear Programming(LP) , Integer Programming (IP) and Binary Integer Programming(BIP).

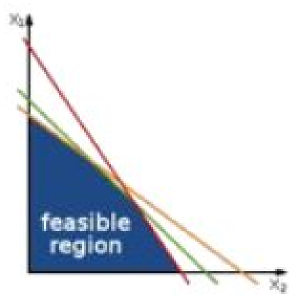

Figure 11: The feasible region, or convex polygon satisfying all constraints in a Linear Program(LP)

Linear Programming (LP)

A LP method is an optimization method that operates on linear constraints and objectives.
Geometrically, this can be interpreted in 2D. The set of constraints on a linear program bounds its decision variables to a convex polygon in the 2D plane. It then tries to find the point inside this convex polygon that maximizes or minimizes the objective function. Typically, for linear programs, this will lie on the edge of the polygon.
Linear programs are expressed as follows:

$$\text{maximize} \quad c^T x$$

$$\text{subject to} \quad Ax \leq b$$

$$\text{and} \quad x \geq 0$$

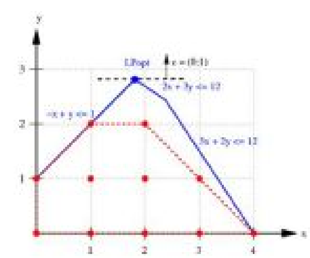

Figure 12: A demonstration of LP-relaxation on an Integer Program

Integer Programming (IP)

Integer Programming (IP) is a special case of Linear Programming(LP) where the decision variables are integral, i.e, the problem comes with an additional constraint:

$$x \in \mathbb{Z}$$

They are typically a harder class of problems than LP because they are NP-hard and cannot be efficiently solved in the worst case. Algorithms used to solve IP include *Cutting-Plane Method*, *Branch and Bound*, *Branch and Cut*, *Branch and Price* and *Delayed Column Generation*. NP-complete problems such as *SAT*, *Travelling Salesman*, *Vertex Cover* and *Set Packing* can all be solved using IP methods.

Binary Integer Programming(BIP)

Binary Integer Programming (BIP) is a special case of Integer Programming (IP) where the decision variables

reinvested into your portfolio. In fantasy cricket, your reward and spending are separate entities. No matter

how much your reward, you will always be restricted to the same budget restrictions as before throughout the tournament.

• Prices of assets do not change Continuing from last point, the price of the asset also doesn't change in fantasy cricket.

• Negative returns are rare Although small negative returns are theoretically possible, they are few and far between.

These key difference force us to adapt our finance model to fit the criteria.

5 Math & Methodology

5.1 Theory

We attempt to formulate our problem as a mathematical optimization problem. Most mathematical optimization problems have 3 key components:

• Decision Variables These variables make the ulti- mate decision of the optimization problem.

• Constraints These are the restrictions imposed upon the decision variables.

• Objective Function It defines what must be opti- mized, in terms of the decision variables.

These 3 components are essential to formulate an opti- mization problem. Particularly relevant to our approach is Linear Programming(LP) , Integer Programming (IP) and Binary Integer Programming(BIP).

Figure 11: The feasible region, or convex polygon satis- fying all constraints in a Linear Program(LP)

Figure 12: A demonstration of LP-relaxation on an Inte- ger Program

Integer Programming (IP)

Integer Programming (IP) is a special case of Linear Pro- gramming(LP) where the decision variables are integral, i.e, the problem comes with an additional constraint:

x 2 Z

They are typically a harder class of problems than LP be- cause they are NP-hard and cannot be efficiently solved in the worst case. Algorithms used to solve IP include Cutting-Plane Method, Branch and Bound, Branch and Cut, Branch and Price and Delayed Column Generation. NP-complete problems such as SAT, Travelling Sales- man, Vertex Cover and Set Packing can all be solved us- ing IP methods.

Binary Integer Programming(BIP)

Binary Integer Programming (BIP) is a special case of Integer Programming (IP) where the decision variables

10

Linear Programming (LP)

A LP method is an optimization method that operates on linear constraints and objectives. Geometrically, this can be interpreted in 2D. The set of constraints on a linear program bounds its decision vari- ables to a convex polygon in the 2D plane. It then tries to find the point inside this convex polygon that maximizes or minimizes the objective function. Typically, for linear programs, this will lie on the edge of the polygon. Linear programs are expressed as follows:

maximize cTx

subject to Ax □ b

and x 0