Higher Education
Australian Mathematical Sciences Institute

Mathematical Models for Cricket Team Selection

Hamish Thorburn
Dr Michael Forbes
The University of Queensland

Australian Government
Department of Education

26/02/2015

1 Abstract

An attempt was made to try and select the Australian Test Cricket Team for then-upcoming series against India in December-January 2014-2015. Data was collected pertaining to 37 Australian cricket players, relating to both recent form (the 12 months prior to commencement of the project) and career form across multiple formats of the game. The team was selected using a mixed-integer-programming (MIP) model, after processing the statistics collected to create usable parameters for the model. It was found that the team selected by the MIP model shared only 5 of the 11 players with the actual team selected by the Australian Board of Selectors to compete in the series. When altering parameters of the model, it was found that the batting diversity of the team could be doubled while only losing 0.008 of the available talent of selected players. The reduced costs were calculated to determine how close unselected players were to being selected, and what they would have to increase their batting/bowling averages to be considered for selection. Finally, we compared the ICC player ratings to our calculated batting and bowling indices, to try to determine the optimal weighting between the different statistics. It was found that batting average was most important in batting performance (but was more important in test matches than one-day matches) and that bowling average and economy were equally important in bowling performance.

2 Introduction

2.1 The Game of Cricket

Cricket is a sport composing of opposing teams of 11 players each side. The teams take turns at batting and bowling. When batting, the objective of the game is to try and score as many 'runs' as possible, while the opposing team (who is bowling at this time) is trying to prevent this. A batting teams 'turn' at batting (called and 'innings') can end 3 ways. The allotted time for the innings can pass (not applicable in all forms of the game), the batting team can 'declare' and simply decide to stop batting, or the field time can dismiss/'get out' 10 of the batting teams players (out of the 11). After both teams have had a turn at batting (called an innings), the team with the most runs is declared the winner.

Cricket in Australia is played in a number of formats. The highest level is international test matches (tests). These matches are played against other countries, and are played over 5 days with each team batting twice. There is no limit on the number of overs or time a team can bat for (aside from the 5 day limit on the game). A similar format, but of a lower level is the Sheffield Shield (SS), a domestic competition within Australia. These games are similar to the tests, but last only 4 days instead of 5, and are played only between Australian States. The final format considered in this report are One-Day Internationals (ODIs). These games are again played against other countries, however, each team only bats once, and is limited to batting to 50 overs. In addition, different cricket balls are used and there are different fielding restrictions, which leads to quicker scoring by the batsmen.

Players on the team often have specialized positions or roles in the team. There are specialist batsman - players who are good at batting and generally are expected to score the bulk of the runs on the team. Also there are specialist bowlers - players who are better are bowling and try to take the majority of the wickets for the team. Within the bowlers there are two different styles. The majority of bowlers are fast/pace bowlers, who bowl the ball quickly and generally take less wickets, but are harder to score runs off. The other type of bowlers are spin bowlers, who get the ball to turn after bouncing off the pitch, and often can take more of the wickets. Some players have similar batting and bowling abilities, and are regarded as 'all-rounders'. And finally, on each team, there is a wicket-keeper (a specialist fielding position, usually also a specialist batsman). It is important in a team to have all of these roles present, as different strategies require different roles to carry out.

Cricket is a popular sport throughout the world, particularly in Australasia, Southern Asia and the United Kingdom. Given the popularity, range and (more recently) the commercialization of the game, much thought has gone into how best to select a cricket team. There are many aspects to team selection to consider - the number of specialist batsman/bowlers in a team, the ratio of right to left handed players in the team, and so on. This factors combine to make optimal team selection a non-trivial problem.

2.2 Background of the Australian Cricket Team

In December-January 2014/2015 the India cricket team toured Australia, playing 4 test matches. The Australian Board of Selectors chose the following players for the squad to face the Indian team

Table 1: The Squad chosen to face India. * indicates player was chosen in the team for the first match

Player	**Role**
Michael Clarke*	Batsman
Brad Haddin*	Batsman/Wicket-Keeper
Ryan Harris*	Bowler
Josh Hazlewood	Bowler
Mitchell Johnson*	Bowler
Nathan Lyon*	Spin Bowler
Mitchell Marsh*	All-Rounder
Shaun Marsh	Batsman
Chris Rogers*	Batsman
Peter Siddle*	Bowler
Steve Smith*	Batsman
David Warner*	Batsman
Shane Watson*	All-Rounder

Table **??** shows the official squad selected to face the Indian cricket team (although Mitchell Starc, a bowler, was added before the second test match, and Joe Burns, a Batsman, was added before the third). The majority of the team was similar to the team that had played against Pakistan a couple of months earlier, and against South Africa in February 2014 [**?**].

3 Problem Formulation and Development

3.1 Data Selection

All data was taken from the *cricinfo* website [?]. The data for 51 Australian Players was collected. These players were selected on the basis of having previously played in the Australian Test Team in the previous 12 months, or showing strong form (being in the top 50 ranked players for either runs scored or wickets taken) in the Sheffield Shield (The Australian Domestic Cricket Competition) either in the current season (2014-2015) or the previous season (2013-2014) - referred to as SS15 and SS14 from here on. Players were then excluded if they had not played at least 20 career Sheffield Shield games, had not played at least either 2 tests or 2 first class games in the selection period or had retired from test cricket/were otherwise ineligible for the Australian Test Cricket team. After removing these players, 37 players remained for consideration (it should be noted that one of the players selected by the Australian Selectors - Mitchell Marsh - was excluded due to the lack of recent matches played).

3.2 Problem Formulation

Using these data sets, we were able to formulate our problem as a Mixed Integer Programming (MIP) model to select our team. We defined the following variables:

$$x_{ij} = \begin{cases} 1 & \text{if Player } j \text{ is selected in the team for role } i \\ 0 & \text{otherwise} \end{cases}$$

We also have the following data:

- $\mathcal{P}$ - the set of all players. The index j will be used as the index for the set of players in this report
- $\mathcal{R}$ - the set of roles in the team. 1 is defined as a batsman, 2 as a bowler, 3 as an all-rounder. The index i will be used as the index for the set of roles in this report
- $\mathcal{F}$ - the set of formats in the team. 1 is defined as test matches, 2 as ODIs, 3 as SS15 matches and 4 as SS14 matches. The index k will be used as the index for the set of formats in this report
- $\mathcal{S}$ - the set of specialist spin bowlers. Note that $\mathcal{S} \subset P$
- $\mathcal{W}$ - the set of specialist wicketkeepers. Note that $\mathcal{W} \subset \mathcal{P}$
- $P_{i,j}$ - the player talent score for player j in role i. Details on how this is calculated are given in section **??**

Our problem was then formulated as follows:

$$\max \sum_{i \in \mathcal{R}, j \in \mathcal{P}} P_{ij} X_{ij} \tag{1}$$

Subject to the following constraints:

$$\sum_{i,j} X_{ij} = 11 \tag{2}$$

$$\sum_{i \in [1,3], j} X_{ij} \geq 7 \tag{3}$$

$$\sum_{i \in [2,3], j} X_{ij} \geq 4 \tag{4}$$

$$\sum_{j \in \mathcal{S}} X_{2j} \geq 1 \tag{5}$$

$$\sum_{j \in \mathcal{K}} X_{1j} \geq 1 \tag{6}$$

$$\sum_{i} X_{ij} \leq 1, \ \forall j \in \mathcal{P} \tag{7}$$

Equation **??** is to ensure that 11 players are selected for the team, equation **??** ensures that at least 7 players are chosen as team batsmen (either specialist batsmen or all-rounders), equation **??** ensures that at least 4 players are chosen as team bowlers (either specialist bowlers or all-rounders), equation **??** forces at least 1 specialist spin bowler to be selected,equation **??** forces at least 1 specialist wicketkeeper to be selected, and equation **??** prevents a player for being selected in more than 1 role for the team.

3.2.1 Player Talent Evaluation

The players were evaluated based on the following methods proposed in [**?**].

Once the eligible players had been selected, data was collected for the following fields (where applicable to each player):

- International Test statistics over the previous 12 months
- Career International Test statistics
- International One-Day statistics over the previous 12 months
- Sheffield Shield statistics for the 2014-2015 season up until November 27 2014 (comprising of 4-5 games)
- Sheffield Shield statistics for the 2013-2014 season
- Career Sheffield Shield statistics

The statistics collected were average (average number of runs per dismissal) and strike-rate (average number of runs scored per 100 balls faced) for batsmen, and average (average number of runs conceded per wicket taken), strike-rate (average number of balls bowled per wicket taken) and economy (average number of runs conceded per 6 balls bowled) for bowlers.

Recent form

The recent form of each player in each format was based on the statistics collected over a 12 month period concluding in late November 2014 (roughly the start of this project). For convenience, the following variables will be defined:

- c_{ijk} as the overall talent index for player j in role i in format k
- $Y_{m,k}$ as batting statistic m (where m takes values of 1 for batting average and 2 for batting strike-rate) of player j in format k
- Z_{njk} as bowling statistic n (where n takes values of 2 for bowling average, 2 for economy and 3 for bowling strike-rate) of player j in format k

For batsmen, a product of the batting statistics was used to define ability. We defined the raw batting indices $U_{1jk} = (Y_{1jk}^{\alpha_k})(Y_{2jk}^{1-\alpha_k})$, where $0 \leq \alpha_k \leq 1$ is a weighting constant for the two statistics. We set $\alpha_k = 1$ for $k = 1, 3, 4$ and $\alpha_k = 0.4$ for $k = 2$ (we chose the same value for α for tests, SS15 and SS14 matches due to these formats being similar). This is calculated for every batsman in every format available. These are then used to calculate the true batting indices:

$$c_{1jk} = \left(\frac{U_{1jk}}{\sum_{j,k} U_{1jk}} \right) \times n_1$$

Where n_1 is the number of player/format statistics are available for batting (so, if there are statistics available for 10 players for tests, 20 players for ODIs, 20 players for the SS15 games and 30 players for the SS14 games, then $n_1 = 10 + 20 + 20 + 30 = 80$)

For bowlers, a similar approach was used, but had to be tweaked due to the fact that lower values for bowling statistics indicate better bowling. We define $U_{2jk} = (Z_{1jk})^{\alpha_{1k}}(Z_{2jk})^{\alpha_{2k}}(Z_{3jk})^{1-\alpha_{1k}-\alpha_{2k}}$, where there values of α_{1k} and α_{2k} are used to weight the different statistics. The different values for these weights are shown in **??**.

Table 2: The different weightings used in different formats

Format	α_1	α_2
Test	0.5	0
ODI	0	0.2
SS14	0.5	0
SS13	0.5	0

Again we chose the same value for α_1 and α_2 for tests, SS15 and SS14 matches due to these formats being similar. Using this we define the bowling indices:

$$V_{jk} = k - \left(\frac{U_{2jk}}{\sum_{jk} U_{2jk}} \right)$$

where k is a chosen value so that all the V_j values are positive. In our case, we chose

$$k = \max_{j,k} \frac{U_{2jk}}{\sum_{jk} U_{2jk}} + 0.01$$

Finally, we define $c_{2jk} = \left(\frac{V_{jk}}{\sum_{j,k} V_{jk}} \right) \times n_2$

where n_2 is defined analogously to n_1.

We used a scale-adjustment procedure to rescale the bowling indices so they had equal variance to the batting indices. This procedure (also used in [?]) iteratively scales the bowling indices, defining the $(p+1)$th bowling indices as

$$c_{2jk}^{p+1} = (c_{2jk}^{p})^{\sigma_{c_1}/\sigma_{c_{2jk}^{p}}}$$

Once this was completed, the all-rounder indices were calculated, using

$$c_{3jk} = (c_{1jk})^{\beta}(c_{2jk})^{1-\beta}$$

with β chosen as a weighting constant, in this case, $\beta = 1/2$.

Finally, for each role, the combined talent for each player was calculated via equation **??**

$$C_{ij} = \frac{\sum_k w_k * c_{ijk}}{\sum_k A_{jk}} \tag{7}$$

Where C_{ij} is the overall recent form of player j in role i, w_k is the weighting of format k (in this project the weightings were assigned as 3,1,2,1 for test, ODI, SS15 and SS14 respectively) and A_{jk} is a binary indicator of whether or not player j had played any cricket in format k over the data collection period.

Career Talent

The above process was repeated using the career statistics in both test and Sheffield shield formats for each player where applicable. The only difference was in the combining of different formats to create the overall career talent for each player. If the player had played ≥ 20 test matches, then only test match statistics were considered in career form. Had the player played between 1 and 20 tests, then a weighted average of test and Sheffield Shield form (weighted on number of test matches played) was used. Had the player not played any test matches previously, then only Sheffield Shield form was used.

Once both recent form and career talent had been calculated, the two scores were averaged to obtain the final overall talent score for each player in each role (P_{ij}).

3.3 Initial Results

The team initially picked is shown in table **??**

It can be seen from table **??** that the selected team is significantly different from the team selected in the first test (shown in table **??**). In fact only 5 out of 11 players were selected by both the selectors and MIP model (Michael Clarke, Steve Smith, David Warner, Mitchell Johnson and Ryan Harris). Key absences include wicket-keeper Brad Haddin (who is also the team vice-captain), and Nathan Lyon as the spin bowler.

Table 3: The initial team selected

Player	**Role**
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Batsman/Wicket-Keeper
Peter Handscomb	Batsman/Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

4 Changing the format weighting

We next wanted to see what the sensitivity of the team was to the weightings given to each format. We define w_k to be the weighting for format k Initially the relative format weightings were set at 3,1,2,1 for test, ODI, SS15 and SS14 (w_1,w_2,w_3 and w_4) respectively. The test weighting was obviously set as the highest, and the SS15 weighting was set higher than ODI and SS14 as it was more recent than SS14, and is a more similar format to test matches than one-day internationals.

4.1 Decreasing the SS15 weighting

As can be seen from the difference between our selected team (table **??**) and the actual selected team (table **??**), it appears that the selectors seem to weight recent test appearances more favorably than current form in the Sheffield Shield. To examine this, we reduced the SS15 weighting (w_3) incrementally from 2 down to 1 and watched the changes in the team.

$w_3 = 2$

Player	Role
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_3 = 1.8$

Player	Role
~~Adam Voges~~ Shaun Marsh	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_3 = 1.4$

Player	Role
Shaun Marsh	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
~~Peter Nevill~~ Shane Watson	~~Wicket-Keeper~~ All-Rounder
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_3 = 1.3$

Player	Role
Shaun Marsh	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
~~Nathan Rimmington~~ Josh Hazlewood	Bowler
Shane Watson	All-Rounder
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

Figure 1: Changes in the team at various values of w_3

The changes in the team, as well as the values of w_3 at which these changes occurred are shown in figure ??. It can be seen that the resulting team more closely matches the selected team. Now six of the eleven players chosen were also chosen for the first test, and two others (Josh Hazlewood and Shaun Marsh) were additional members of the squad. Indeed, both these players were selected to play the second test (but in fairness, Shaun Marsh was only selected to cover for an injured Michael Clarke). It can be seen that Brad Haddin, Nathan Lyon, Peter Siddle or Chris Rogers aren't being selected at this stage.

4.2 Decreasing the SS14 weighting

We now reduce the weighting for SS14 matches from 1 down to 0, as this is the oldest data used, and will be the first to become irrelevant.

$w_4 = 1$

Player	Role
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_4 = 0.9$

Player	Role
~~Adam Voges~~ Shaun Marsh	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_4 = 0.3$

Player	Role
Shaun Marsh	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
~~Nathan Rimmington~~ Mitchell Starc	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_4 = 0.2$

Player	Role
Shaun Marsh	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
~~Michael Clarke~~ Shane Watson	~~Batsman~~ All-Rounder
Mitchell Starc	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

Figure 2: Changes in the team at various values of w_4

We see the changes in the team at reduced values of w_4 in figure **??**. Again, we do see some players both selected in the initial team and in the extended squad coming into the team, but we also see Michael Clarke (who was widely regarded as one of Australia's top batsmen at the time) dropping out when the weighting for SS14 matches in lowered enough. Again, we see Adam Voges is dropped from the team at only a small reduction in the SS14 weighting.

4.3 Decreasing the ODI weighting

We also looked at the effect of reducing the weighting from one-day games, as this format is the most different to test cricket of the formats considered.

$w_2 = 1$

Player	Role
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_2 = 0.6$

Player	Role
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
~~Steve Smith~~ Shaun Marsh	Batsman

Figure 3: Changes in the team at various values of w_2

The changes to the team can be seen in figure **??**. In this instance, we only see one change, with Shaun Marsh coming in for Steve Smith, meaning that this team also looks very dissimilar to the team selected for the first test.

4.4 Decreasing the test weighting

Finally, we looked at what would happen if the weighting of test matches was reduced.

$w_1 = 3$

Player	Role
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

$w_1 = 2.6$

Player	Role
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
Mitchell Johnson	Bowler
~~Michael Clarke~~ Callum Ferguson	Batsman
Nathan Rimmington	Bowler
Peter Nevill	Wicket-Keeper
Peter Hanscomb	Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

Figure 4: Changes in the team at various values of w_1

As seen in figure **??** the only change occurs at $w_1 = 2.6$, at which point Callum Ferguson comes in for Michael Clarke.

5 The affect of adding batting diversity

While picking talented players is no doubt important, there is some value in having players with a diverse range of batting styles - that is have both right-handed and left-handed batsmen in the team [**?**] (note that since all players on the team are required to bat, in this context, batsmen refer to all players on the team when batting). We looked at adding value to diversity in the batting line-up to our objective function to see how the team is changed, as well as looking at how sensitive the team is to changes.

In order to do this, we need to make some changes to our objective function and constraints. First, we define the following additional data:

- $Right_j = \begin{cases} 1 & \text{if player } j \text{ bats right-handed} \\ 0 & \text{otherwise} \end{cases}$

And we also introduce the following variable:

- $R_l, = \begin{cases} 1 & \text{if } l \text{ players on the team bat right-handed} \\ 0 & \text{otherwise} \end{cases}$

As per [?], if r is the number of right-handed batsmen in the team, the scores for diversity (d) were initially assigned to be

$$d = r(11 - r)$$

To implement this, we define the data

$$d_l = l(11 - l)$$

In addition, we wanted to gradually increase the value of the diversity score of the team, so as to see how the team would change. To do this, we added a diversity score parameter D_v which ranged from 0 to 0.1 in 0.01 increments. The diversity score in the objective function was multiplied by this score. The reason the diversity parameter is so small is because the maximum possible diversity score is $6 * 5 = 30$. To put this in perspective, the objective value for our initial team was approximately 14.8. Therefore, if we included this diversity score without scaling it back, very quickly diversity would dominate the objective function, and all difference from player talent would be diminished.

This is implemented in our MIP model by altering the objective function (equation **??**) to be

$$\max \sum_{i \in Roles, j \in Players} P_{i,j} X_{i,j} + D_v \sum^{1} 1_{l=0} d_l R_l \tag{8}$$

as well as adding the additional constraints

$$\sum_{l=1}^{11} R_l = 1 \tag{9}$$

$$\sum_{i,j} Right_j X_{i,j} = \sum_{l=1}^{11} l R_l \tag{10}$$

Equations **??**and **??** are to ensure that $R_l = 1$ if and only if l is equal to the number of right-handed batsmen in the team.

When these changes are made to the MIP model, we obtain the following teams

V=0

Player	Role	Hand
Adam Voges	Batsman	R
David Warner	Batsman	L
David Hussey	Batsman	R
Mitchell Johnson	Bowler	L
Michael Clarke	Batsman	R
Nathan Rimmington	Bowler	R
Peter Nevill	Wicket-Keeper	R
Peter Hanscomb	Wicket-Keeper	R
Ryan Harris	Bowler	R
Steve O'Keefe	Spin Bowler	R
Steve Smith	Batsman	R

V=0.01

Player	Role	Hand
~~Adam Voges~~ Shaun Marsh	Batsman	~~R~~ L
David Warner	Batsman	L
David Hussey	Batsman	R
Mitchell Johnson	Bowler	L
Michael Clarke	Batsman	R
Nathan Rimmington	Bowler	R
Peter Nevill	Wicket-Keeper	R
Peter Hanscomb	Wicket-Keeper	R
Ryan Harris	Bowler	R
Steve O'Keefe	Spin Bowler	R
Steve Smith	Batsman	R

V=0.04

Player	Role	Hand
Shaun Marsh	Batsman	L
David Warner	Batsman	L
David Hussey	Batsman	R
Mitchell Johnson	Bowler	L
Michael Clarke	Batsman	R
~~Nathan Rimmington~~ Mitchell Starc	Bowler	~~R~~ L
Peter Nevill	Wicket-Keeper	R
Peter Hanscomb	Wicket-Keeper	R
Ryan Harris	Bowler	R
Steve O'Keefe	Spin Bowler	R
Steve Smith	Batsman	R

Figure 5: Changes in the team at various values of D_v

As can be seen from figure **??**, two changes are made to bring in left-handed players at the expense of right-handed players. Adam Voges is once again dropped for Shaun Marsh, and Mitchell Starc again comes in for Nathan Rimmington. However, it should be noticed that for $D_v \in [0.04, 0.1]$, no other changes were made, even though the diversity score could have been improved. This shows that there is a clear gap between the talent of the remaining right-handed players in the team, and the left-handed players that could potentially be selected.

What should also be considered is how much talent is lost when picking players for diversity. To do this, we take the sum of the talent of all players selected in the team, and look at how it changes when increasing D_v.

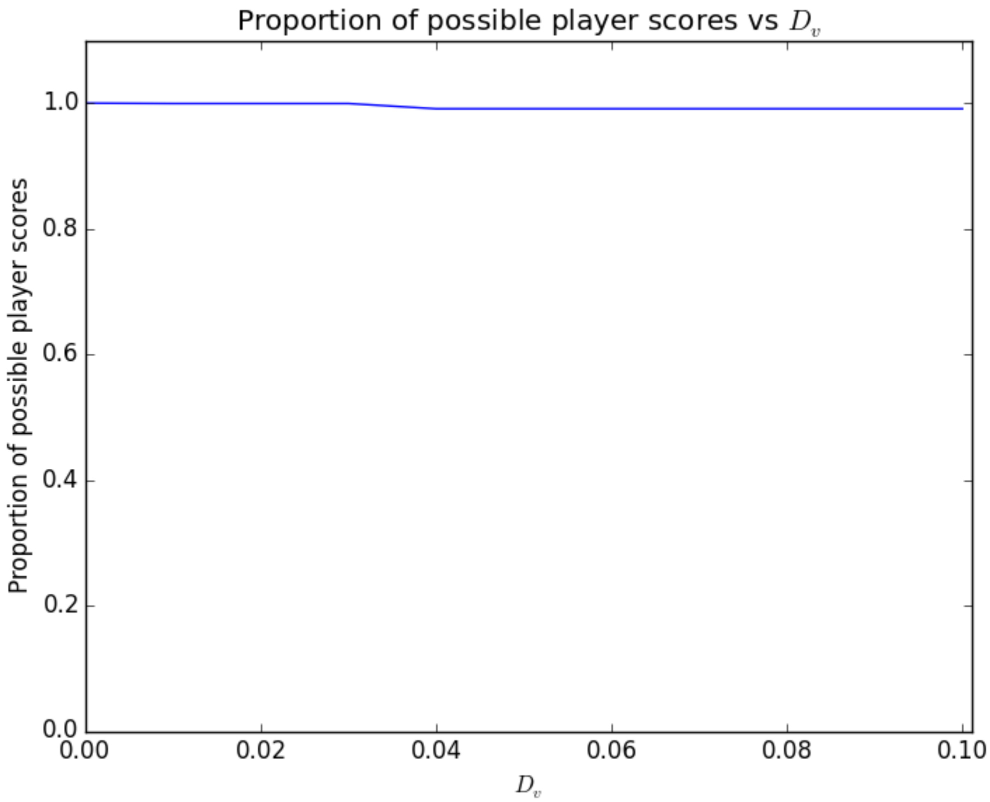

Figure 6: Talent lost when selecting for diversity

Figure **??** shows the talent lost as a proportion of possible talent. The talent of the team at each stage is represented as a proportion of the talent in the team when $D_v = 0$, i.e., when the only term in the objective function is player talent (effective the highest possible talent score for a team). It can be seen that the talent proportion of the team only drops slightly (to a minimum of 0.991) when selecting for diversity.

6 Reduced costs and player selection requirements

After examining the effect of changing coefficients and parameters in our model, we saw that often it was similar players leaving/entering the team, indicating that there are some players on the cusp of selection. This begs the question - how close are these players?

By using reduced costs obtained from the MIP, we can answer the question of how close each non-selected player is to being selected, and even what his average would need to be in a certain format to warrant selection.

This could be done as the LP relaxation of the problem would naturally give an integer solution in every case. Because of this, the integer constraints could be relaxed, so the reduced costs of each variable could be calculated (this can only be done with linear programming models). Using the reduced costs, we were able to work out what the overall talent score of each unselected player would need to be to warrant selection, and (by inverting the process of calculating the batting/bowling indices), we were able to calculate what the players required average would need to be in a given format.

Table **??** shows the reduced cost, current average in a given format and the required average in that format to be selected for the test team. The format chosen was the format in which the player achieved the best average for their role in the past 12 months. In can be seen that some players (such as Shaun Marsh, Callum Ferguson and Joe Burns) are right on the cusp of selection (indeed, Burns was called into the squad as an injury cover for the 3rd test in Melbourne [**?**]).

What else is interesting to see is how far away some of the currently 'official' selected players are from being selected by the MIP model. For example, Nathan Lyon, with a reduced cost of 0.5 (with a recent test bowling average of approximately 47) would need to improve his bowling average in tests to approximately 0.7 - a task that is virtually impossible. Comparing this to Fawad Ahmed (another spin bowler) for example, with a reduced cost of 0.07, and one can see the vast difference between the Australian cricket selectors and selecting players using a MIP model.

Player	Role	Reduced Cost	Format	Current Average	Required Average
Ashton Agar	Bowler	0.67	SS15	26.33	0.26
Alex Doolan	Batsman	0.44	Test	35.72	70.16
Aarron Finch	Batsman	0.43	ODI	35.25	118.59
Brad Haddin	Batsman	0.25	Test	29.78	49.64
Chris Hartley	Batsman	0.42	SS15	38.20	71.41
Callum Ferguson	Batsman	0.03	SS15	70.40	72.42
Chadd Sayers	Bowler	0.10	SS15	25.54	16.96
Chris Rogers	Batsman	0.30	Test	107.00	71.03
Ed Cowan	Batsman	0.38	SS15	40.71	70.75
Fawad Ahmed	Bowler	0.07	SS15	30.37	23.50
George Bailey	Batsman	0.37	Test	36.66	68.75
Glenn Maxwell	Batsman	0.38	SS14	45.33	101.08
Joe Burns	Batsman	0.07	SS15	58.83	64.47
Joe Mennie	Bowler	0.37	SS14	32.00	0.34
James Faulkner	Bowler	0.18	SS15	28.00	9.37
Josh Hazlewood	Bowler	0.16	SS14	19.77	3.92
James Hopes	Bowler	0.11	SS15	17.00	10.16
Jonathon Wells	Batsman	0.35	SS15	44.83	58.60
Luke Feldman	Bowler	0.07	SS14	23.75	13.01
Michael Klinger	Batsman	0.28	SS15	44.00	66.48
Mitchell Starc	Bowler	0.12	SS15	20.14	11.30
Michael Hogan	Bowler	0.19	SS14	25.86	4.59
Mark Cosgrove	Batsman	0.26	SS15	36.20	56.86
Marcus Harris	Batsman	0.66	SS15	16.16	42.42
Nathan Lyon	Bowler	0.50	Test	47.26	0.73
Peter Forrest	Batsman	0.29	SS14	68.58	108.70
Peter Siddle	Bowler	0.25	Test	12.14	16.11
Ryan Carters	Batsman	0.41	SS14	53.81	110.32
Rob Quiney	Batsman	0.35	SS14	48.57	99.83
Shaun Marsh	Batsman	0.01	SS15	76.25	77.35
Sam Whiteman	Batsman	0.43	SS14	45.80	107.85
Scott Henry	Batsman	0.37	SS15	52.66	80.85
Shane Watson	Batsman	0.11	Test	42.44	51.50
Tim Paine	Batsman	0.58	SS14	31.53	115.60
Tom Cooper	Batsman	0.10	SS15	56.80	65.03
Travis Head	Batsman	0.49	SS14	36.00	107.79
Xavier Doherty	Bowler	0.75	SS14	29.66	16.70

Table 4: Reduced Costs and Required Averages for non-selected Players

7 Comparing parameters to player rankings

Almost all parameters used in this study have been chosen arbitrarily. While in most cases, this has been done as there is no objective criteria to measure our choices against, we can use the ICC player ratings to determine if the correct weightings for α_k, α_{1k} and α_{2k} have been chosen. We can do this by computing the batting and bowling indices at different levels of each parameter, regressing these against the ICC rating for the given player, and statistically analysing the result.

The ICC player ratings are a system in which players are rated based on their recent performances, and then ranked accordingly [**?**]. It is regarded as a definitive ranking of players in the wider cricket community. The players are given a rating between 0 and 1000 based on recent performance, and take into account the quality of the opposition faced, as well as the achievement of the player. The players are then ranked according the their ratings. The ICC website states that 'Over 900 points is a supreme achievement... 500 plus is a good, solid rating'. The ratings exist for both test and ODIs.

7.1 Batting Performances and Ratings

Since the ratings were bounded between 0 and 1000, we needed to initially transform the player ratings onto a continuous scale. This was done using a logistic function, similar to a logistic regression. If we let Rt_j be the rating of player j and T_j be the transformed rating of player j, then the transformation is

$$T_j = log(\frac{\frac{Rt_j}{1000}}{1 - \frac{Rt_j}{1000}}) \tag{11}$$

The transform in equation **??** is useful as it both preserves relative ordering of ratings, as well as mapping them from the interval $[0, 1000]$ to $\mathbb{R}$.

We performed this transform, and then plotted the resulting transformed ratings against the raw batting indices (that is the batting indices before being scaled back by the total sum of the indices and number of scores present) calculated using different values of α. This was performed for both test and ODI ratings

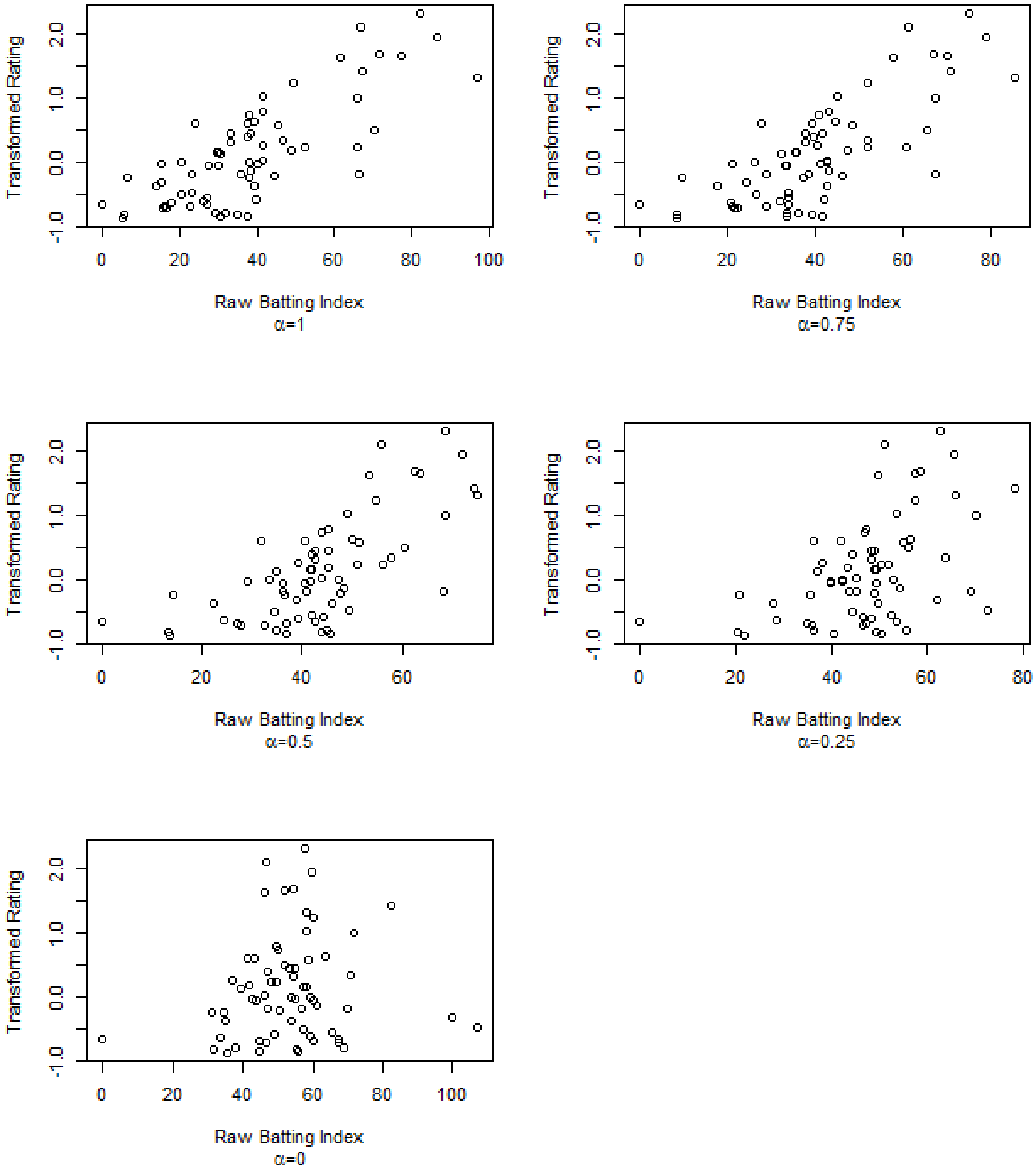

Figure 7: Transformed Test Ratings vs Raw Batting index for varying levels of α

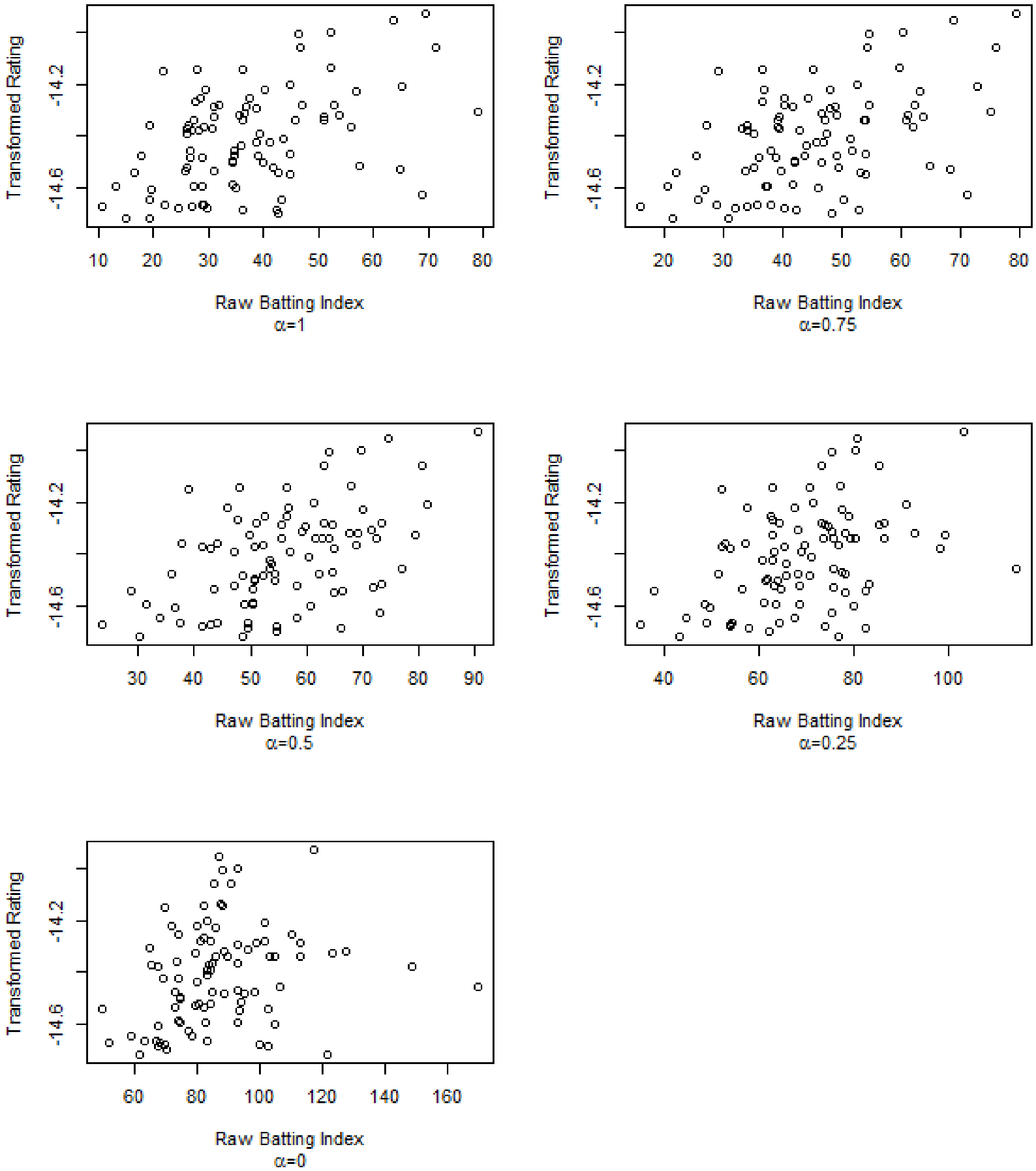

Figure 8: Transformed Test Ratings vs Raw Batting index for varying levels of α

As can be seen from figures **??** and **??**, for higher values of α (that is, weighting the batting average more than the strike rate), there appears to be a more true linear relationship between Transformed rating and raw batting index. This relationship appears to be reasonably strong for test performances, however the relationship is still weak, even for the higher values of α for the ODI performances. To confirm this, a linear regression was performed between the transformed ratings and the raw batting index calculated using differing values of α between 0 and 1 for both test and ODI players. The R^2 value was found to quantify the strength of the relationship, and the p-value of the correlation was also found to check for significance.

Table 5: Strength of relationship between transformed ratings and raw batting indices for differing levels of α

α	Test R^2	Test p-value	ODI R^2	ODI p-value
0.00	0.01	0.32	0.05	0.04
0.10	0.08	0.02	0.09	¡0.01
0.20	0.18	¡0.01	0.14	¡0.01
0.30	0.30	¡0.01	0.19	¡0.01
0.40	0.40	¡0.01	0.22	¡0.01
0.50	0.47	¡0.01	0.24	¡0.01
0.60	0.53	¡0.01	0.25	¡0.01
0.70	0.56	¡0.01	0.25	¡0.01
0.80	0.58	¡0.01	0.25	¡0.01
0.90	0.60	¡0.01	0.24	¡0.01
1.00	0.61	¡0.01	0.23	¡0.01

Table **??** appears to back up what is shown in figures **??** and **??**. As α_1 is increased for test performances, the correlation between the raw batting index and the player ratings becomes stronger (and the p-value lowers). With the ODI performances, we see that the relationships are generally weaker, and are maximized at around $\alpha_2 \in [0.6, 0.8]$. However, even though the correlation is deemed to be significant at a 0.05 level for the ODI ratings, the plots show relationships which may not be linear (particularly for lower levels of α_2), so caution needs to be made when drawing conclusions about these relationships.

7.2 Bowling Performances and Ratings

The ratings were again transformed as per equation **??**. The α_1 (weighting of bowling average) and α_2 (weighting of bowling economy) were varied between 0 and 1 in increments of 0.25, and the raw bowling indices were calculated and plotted against the transformed ratings.

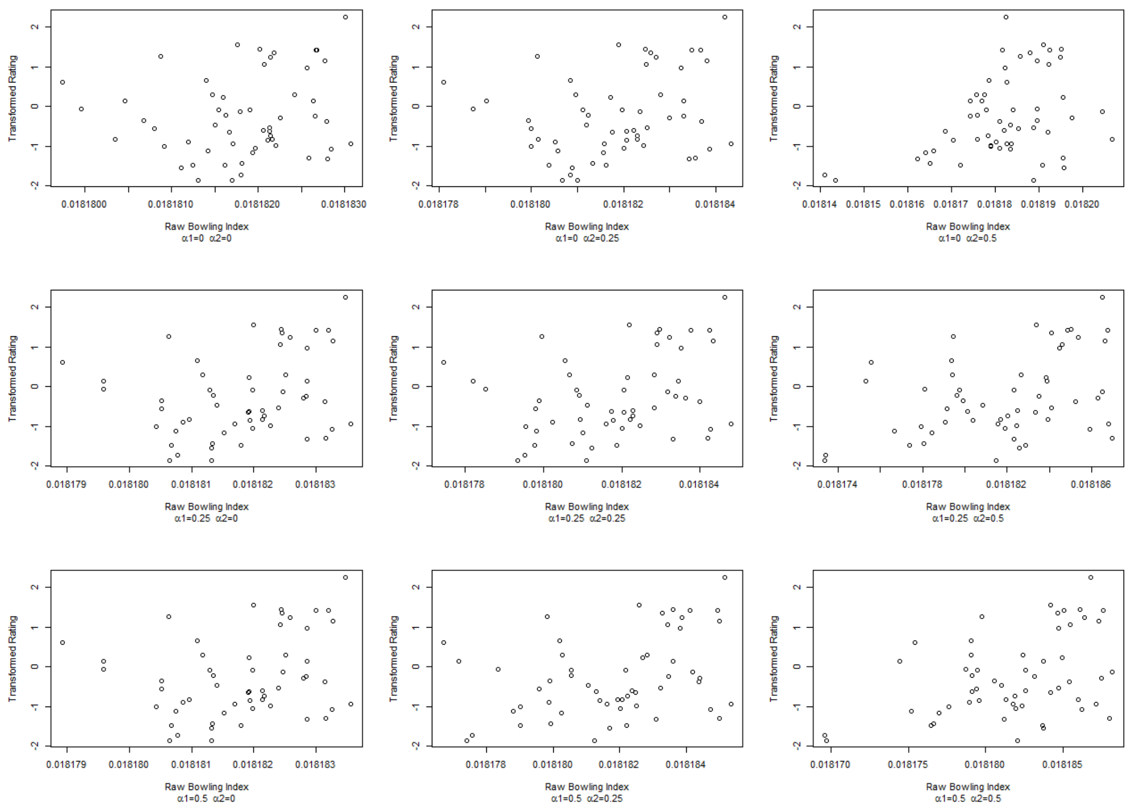

Figure 9: Transformed Test Ratings vs Raw Bowling index for varying levels of α_1 and α_2

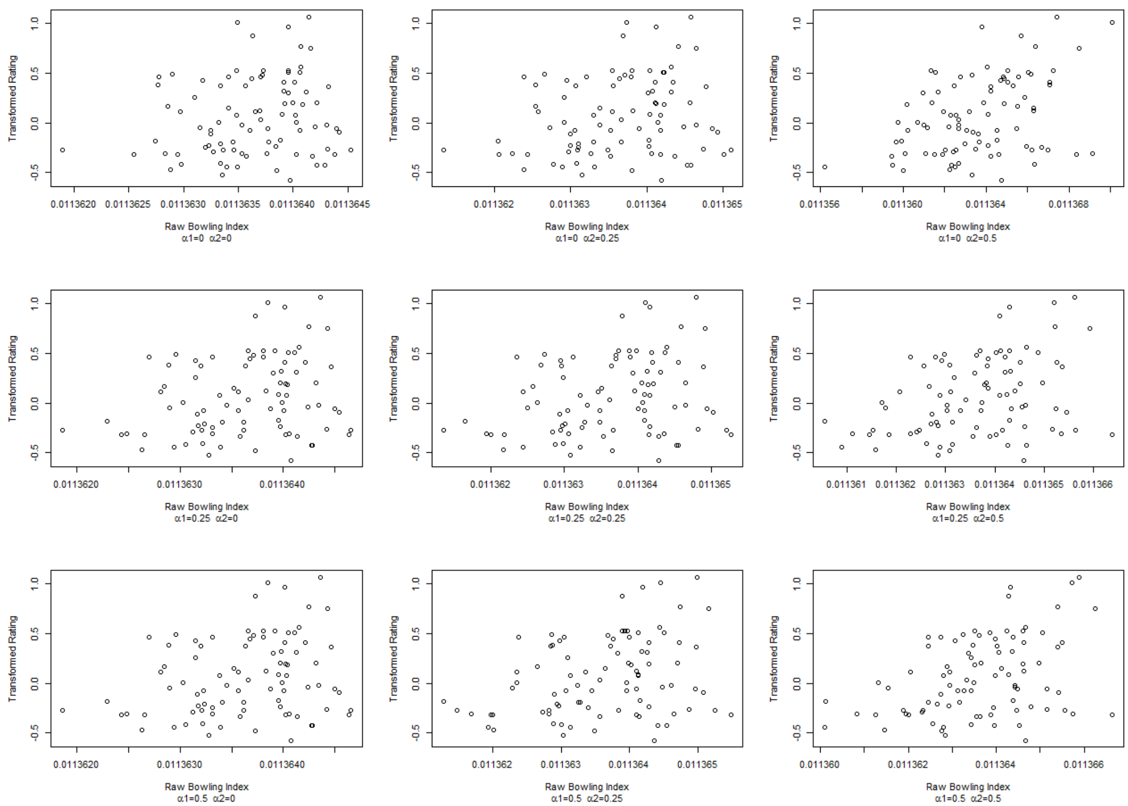

Figure 10: Transformed ODI Ratings vs Raw Bowling index for varying levels of α_1 and α_2

Looking at figures **??** and **??**, it can be seen that the relationships between raw bowling indices and transformed ratings is much less clear for bowlers than batsmen. It does appear that increasing both α_1 (the weighting for bowling average) and α_2 (the weighting for bowling economy) increases the strength of the relationship in both tests and ODIs - meaning that bowling strike-rate is less important. Whilst this is intuitive in ODIs, it is surprising that the strike-rate is not valuable in tests, as there is no time limit on a batting innings, so it is valuable to be taking wickets often.

Again, we looked at R^2 values to determine the strength of relationships between bowling indices and transformed ratings.

Table 6: The R^2 values for the association between transformed ratings and test bowling indices for varying α_1 and α_2

	α_2					
α_1	0	0.1	0.2	0.3	0.4	0.5
0	0.02	0.03	0.05	0.07	0.09	0.13
0.1	0.03	0.04	0.06	0.08	0.11	0.14
0.2	0.04	0.05	0.07	0.10	0.12	0.15
0.3	0.05	0.06	0.08	0.11	0.14	0.16
0.4	0.06	0.07	0.10	0.12	0.15	0.17
0.5	0.07	0.09	0.11	0.13	0.16	0.17

Table 7: The p-values for the association between transformed ratings and test bowling indices for varying α_1 and α_2

	α_2					
α_1	0	0.1	0.2	0.3	0.4	0.5
0	0.26	0.18	0.11	0.06	0.02	0.01
0.1	0.20	0.13	0.07	0.04	0.01	¡0.01
0.2	0.15	0.09	0.05	0.02	0.01	¡0.01
0.3	0.11	0.06	0.03	0.01	0.01	¡0.01
0.4	0.08	0.04	0.02	0.01	¡0.01	¡0.01
0.5	0.06	0.03	0.01	0.01	¡0.01	¡0.01

Table 8: The R^2-values for the association between transformed ratings and ODI bowling indices for varying α_1 and α_2

	α_2					
α_1	0	0.1	0.2	0.3	0.4	0.5
0	0.02	0.03	0.04	0.05	0.07	0.09
0.1	0.03	0.04	0.05	0.06	0.08	0.10
0.2	0.04	0.04	0.06	0.07	0.09	0.12
0.3	0.04	0.05	0.07	0.08	0.10	0.13
0.4	0.05	0.06	0.07	0.09	0.12	0.14
0.5	0.06	0.07	0.08	0.10	0.13	0.15

Table 9: The p-values for the association between transformed ratings and ODI bowling indices for varying α_1 and α_2

	α_2					
α_1	0	0.1	0.2	0.3	0.4	0.5
0	0.15	0.11	0.07	0.04	0.02	¡0.01
0.1	0.11	0.07	0.04	0.02	0.01	¡0.01
0.2	0.07	0.05	0.03	0.01	¡0.01	¡0.01
0.3	0.05	0.03	0.02	0.01	¡0.01	¡0.01
0.4	0.04	0.02	0.01	¡0.01	¡0.01	¡0.01
0.5	0.02	0.01	0.01	¡0.01	¡0.01	¡0.01

Tables **??** and **??** show that for both formats, the strongest relationship occurs when both $\alpha_1 = 0.5$ (weighting for the bowling average) and $\alpha_2 = 0.5$ (weighting for the economy rate), meaning that the bowling strike rate is not used. This confirms what was seen in the plots. The R^2 values also confirm that these associations are quite weak (but statistically significant, as seen in tables **??** and **??**).

Based on this analysis, we conclude that the appropriate levels for α are 1 for test, SS15 and SS14 and 0.6 for ODI, and the appropriate levels for α_1 and α_2 are 0.5 for all formats.

7.3 Updating the team and the reduced costs

Finally, using these adjusted parameters, the team can be re-selected and the reduced costs re-calculated.

Table **??** shows the team selected with the updated player talent parameters. It can be seen that the only change is James Hopes coming in as a bowler for Nathan Rimmington.

We also re-calculated the reduced costs. Table **??** shows the re-calculated reduced costs and required averages. Again, it shows the average required for the format the player was performing the best in (with the except of players who had played a test match in the past 12 months, for which required test averages are shown). Note that the bowlers in general now need much more reasonable averages than using the previous values for α_1 and α_2 (For example, Nathan Lyon now needs to achieve an average of approximately 7 as opposed to 0.7).

Table 10: The new team. Bold indicates a change from the initial team selected

Player	**Role**
Adam Voges	Batsman
David Warner	Batsman
David Hussey	Batsman
James Hopes	**Bowler**
Mitchell Johnson	Bowler
Michael Clarke	Batsman
Peter Nevill	Batsman/Wicket-Keeper
Peter Handscomb	Batsman/Wicket-Keeper
Ryan Harris	Bowler
Steve O'Keefe	Spin Bowler
Steve Smith	Batsman

8 Discussion

What has been shown here is simply the choices that can be made with regards to team selection using a mixed integer programming model, and the resulting consequences. One of the particularly noteworthy aspects is that not matter what parameters one sets with regards to team diversity or format weighting is that the team is consistently different to the team chosen by the Australian Board of Selectors.

However, while we have shown that using a MIP model would select a different team, we have NOT shown that the team selected by the MIP model would in fact perform better than the team chosen by the selectors. It's worth noting that many of the test players selected (who have poor statistics) are playing against far superior opposition than their counterparts in the Sheffield Shield, and indeed because they are playing internationally, do not get the chance to play Sheffield Shield in addition to boost their selection claims. In addition, it should be noted that the team chosen by the selectors won the four match series 2-0, with many of the selected players playing well. In particular, Nathan Lyon (who we used as an example of someone who would not be selected using the MIP model) was the leading wicket-taker in the series, taking 23 wickets at an average of 34.52, and was award man-of-the-match for the first test (the first match the team played after our data collection). So, we cannot actually say that the selectors chose the wrong team, or that the MIP model would have chosen a different one. Realistically, all that can be concluded is that the Australian Cricket Selectors look at more than just player statistics when selecting the team.

Another problem with our MIP model model is the many of the parameters have been arbitrarily chosen. While often there is a fair amount of intuition behind the decisions (e.g., weighting recent test performances more than ODI performances, as test performance is more relevant to selecting a test team), we still haven't shown that these values are correct. This is because the only outcome we can obtain by changing the parameters is to obtain a different team, and again, we have no way of truly comparing one team to another. In the only instance in which could valid our parameters (comparing batting and bowling indices to player ratings), we found that while we were using appropriate values for the batting indices, we were under-weighting bowling economy in test match bowlers. Upon improving this, it was found that a change is made in our selected test team (James

Hopes coming in for Nathan Rimmington). However, the relationships we saw when evaluating the appropriate levels for α, α_1 and α_2 were quite weak, so caution should be made when drawing conclusions from them.

Future directions this research could take would be to try and gauge how well a team selected by this method would actually perform. This could possibly done by looking at historical records of team performance, and calculating the talents scores for the teams retrospectively, then looking at the relationship between team talent scores and performance.

Player	Role	Reduced Cost	Format	Current Average	Required Average
Ashton Agar	Bowler	0.70	SS15	26.33	6.00
Alex Doolan	Batsman	0.44	Test	23.87	70.08
Aaron Finch	Batsman	0.43	ODI	35.25	118.40
Brad Haddin	Batsman	0.25	Test	29.78	49.66
Chris Hartley	Batsman	0.42	SS15	38.20	71.44
Callum Ferguson	Batsman	0.03	SS15	70.40	72.36
Chadd Sayers	Bowler	0.09	SS15	25.54	21.28
Chris Rogers	Batsman	0.30	Test	39.72	71.06
Ed Cowan	Batsman	0.38	SS15	40.71	70.77
Fawad Ahmed	Bowler	0.20	SS15	30.37	20.05
George Bailey	Batsman	0.37	Test	29.20	68.61
Glenn Maxwell	Batsman	0.38	SS14	45.33	101.51
Joe Burns	Batsman	0.07	SS15	58.83	64.40
Joe Mennie	Bowler	0.44	SS14	32.00	5.01
James Faulkner	Batsman	0.27	SS15	62.50	92.66
Josh Hazlewood	Bowler	0.32	SS14	19.77	4.96
Jonothan Wells	Batsman	0.35	SS15	44.83	58.55
Luke Feldman	Bowler	0.20	SS14	23.75	10.19
Michael Klinger	Batsman	0.28	SS15	44.00	66.29
Mitchell Starc	Bowler	0.29	SS15	20.14	10.98
Michael Hogan	Bowler	0.16	SS14	25.86	13.06
Mark Cosgrove	Batsman	0.26	SS15	36.20	56.84
Marcus Harris	Batsman	0.66	SS15	16.16	42.42
Nathan Rimmington	Bowler	0.03	SS15	15.81	14.78
Nathan Lyon	Bowler	0.53	Test	47.26	7.20
Peter Forrest	Batsman	0.29	SS14	68.58	109.42
Peter Siddle	Bowler	0.26	Test	39.38	23.18
Ryan Carters	Batsman	0.41	SS14	53.81	111.22
Rob Quiney	Batsman	0.35	SS14	48.57	100.22
Shaun Marsh	Batsman	0.01	SS15	76.25	77.30
Sam Whiteman	Batsman	0.43	SS14	45.80	108.60
Scott Henry	Batsman	0.37	SS15	52.66	80.92
Shane Watson	Batsman	0.11	Test	42.44	51.50
Tim Paine	Batsman	0.58	SS14	31.53	115.47
Tom Cooper	Batsman	0.10	SS15	56.80	64.93
Travis Head	Batsman	0.49	SS14	36.00	108.49
Xavier Doherty	Bowler	0.79	SS14	29.66	0.00

Table 11: The re-calculated reduced costs and required averages

References

[1] Sharp, GD, Brettenny, WJ, Gonsalves, JW, Lourens, M and Stretch, RA, 2011 *Integer optimisation for the selection of a Twenty20 cricket team* Journal of the Operational Research Society, Volume 62.

[2] S. Siva Sathya, Md. Shad Jamal, 2009 *Applying Genetic Algorithm to Select and Optimal Cricket Team* International Conference on Advance in Computing, Communication and Control.

[3] Whitley, Darrell, 1994. *A genetic algorithm tutorial* Statistics and Computing, Volume 4, Issue 2.

[4] Statsguru, Cricinfo `http://stats.espncricinfo.com/ci/engine/stats/index.html`

[5] International Cricket Council, "About Reliance ICC Cricket", updated 2015, *http://www.icc-cricket.com/player-rankings/about*