How do basketball teams win championships?  
A quantitative analysis on factors determining wins

Relatore
Ch. Prof. Paola Ferretti

Laureando
Matteo Aggio
Matricola 828304

Anno Accademico
2013 / 2014
The sea’s only gifts are harsh blows, and occasionally the chance to feel strong. Now, I don’t know much about the sea, but I do know that that’s the way it is here; and I also know how important it is in life not necessarily to be strong but to feel strong, to measure yourself at least once, to find yourself at least once in the most ancient of human conditions, facing the blind deaf stone alone, with nothing to help you but your hands and your own head.

Primo Levi
Abstract

Since its appearance in the late fifties, evaluation of sport performance represents a growing field of research, and quantitative analysis has established as a useful and efficient instrument in order to assess and evaluate performance of both teams and players. Numerous sports are subject of analysis (baseball, basketball, football, hockey etc.), as well as numerous techniques have been developed to study them, creating a significant body of literature. For instance, analysis of performance within basketball has been carried on from different perspectives, implementing various ways of thinking, which are different both in the subject of evaluation and in the method these evaluations are driven. This research introduce a new approach of team assessment through the use of traditional instruments. In this context, Dean Oliver’s work on measuring performance in the NBA represents a useful method to evaluate a team’s strengths and weaknesses and to analyze its efficiency, while Rough Set Theory is used as innovative approach for basketball analysis. The aim of this research is to quantitatively analyze which are the factors that led to success (or failure) some Italian basketball teams during specific seasons of Legabasket (from 2009/2010 to 2013/2014), in order to obtain certain rules, known as decision rules. Given the cause-effect relationship of the factors involved, a decision table will be set up, presenting a simple method of drawing consistent conclusions from the statistics collected on court and giving explanation of obtained results.
Contents

Abstract 5

1 Introduction 9

2 Literature Review 15
  2.1 The Scoring Models . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
  2.2 The Plus-Minus Model . . . . . . . . . . . . . . . . . . . . . . . . . . 19
  2.3 The Win Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

3 The Model 27
  3.1 The database . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
  3.2 Oliver’s eight factors . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
  3.3 Synergy and Overall indexes . . . . . . . . . . . . . . . . . . . . . . . 33
  3.4 Rough Set Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

4 Results 41
  4.1 Coverage and certainty . . . . . . . . . . . . . . . . . . . . . . . . . . 41
  4.2 (False) Basketball myths . . . . . . . . . . . . . . . . . . . . . . . . . 42
  4.3 General rules . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
  4.4 Strategies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55

5 Conclusions 57

6 Acknowledgements 59

References 61

7 Appendix 63
1 Introduction

It’s about getting things down to one number. Using the stats the way we read them, we’ll find value in players that no one else can see.

The opening reference of this paper is cited in a very popular movie released in 2011, *Moneyball*. Starring Brad Pitt and Philip Seymour Hoffman, the movie takes the audience through the incredible story of a baseball team, the Oakland Athletics, which in the 2002 MLB season managed to become a very competitive team despite facing an unfavorable financial situation, and to hit a successful 20-consecutive wins streak.

The reason why this movie became so popular (it was even nominated best motion picture for the Golden Globe Awards in 2012), omitting its cinematographic characteristics, is that it is focused on a peculiar way of looking at sports, that of quantitative analysis. General Manager of the team and his assistant approach the performance analysis of their players and other potential prospects through the use of *sabermetric*, a so called instrument with the aim of searching for objective knowledge and analysis of baseball.

Introduced in the middle of the 20th century, *sabermetric* probably represented the first attempt to give structure and organization to the world of sport statistical analysis, especially in the field of baseball. In recent years, sport and performance analysis have experienced a huge and ongoing success that has been carried on by the dialectic relation between sports and numbers. Numbers represent a simple and efficient way to understand and explain almost every sport performance, so it was natural for statisticians to study a way to better comprehend how that singular performance was reached, and eventually to forecast a future one.

The reasons why the dialog between sport and numbers has gone so far are essentially three: pure analysis purposes, economic reasons and betting motives. The first two subjects are strictly related to each other - very often researches on performance analysis have an economic purpose - while the third one represents a different system.

Researches within the field of sport analysis are driven by desire to better describe the goals and the structure of the game, and these needs come from different subjects. First of all, coaches, scouts and players. Quantitative sport analysis is not aimed to introduce new play-
ing schemes or training regimens, but to help coaches in their decisions through several statistical measures; for example, by providing formulas for evaluating a team and the contribution of individuals to team success. It can solve some typical problems that coaches face with their team, for instance how does teammate interaction matter, or how it is possible to predict the best lineup against a particular opponent, or how it is possible to know if a single player is actually efficient or not, providing insights for making the coaching mission more effective.

Secondly, numbers and statistics are particularly appreciated by fans. They tend to approach sports discussions mentioning the number of goals their team did in the last game, the length of a winning strike, the efficiency of the interaction between teammates or even the average possession of the ball percentage. Indeed, fans tend to establish “heroes”, players whose technical and athletics qualities are recognized and that tend to carry their teams into a different game level. Quantitative analysis is able to judge objectively these players, and see whether or not their reputation is deserved. Or even spot players who get much less attention then they deserve (given that at high-level sports attention eventually means money, it is clear that it is not just a question of respect). A typical example is given by the NBA MVP (Most Valuable Player) title: an award that attempts to reflect a player contribution to his teams winning percentage, and definitely numbers and quantitative analysis do matter on the process of making this selection.

Another class of subjects interested about this topic is given by contract negotiators, who represent the edge between sport analysis and economic value. Having good statistics is part of the process of understanding the quality of a player, and understanding the quality of a player is part of the process of making a contract. The value of a player is the substantive core of making a deal with him, and signing a mediocre player to a top-dollar contract has some effective issues, for example it can restrict signing other players and maybe it can force a coach to make that player a leader, causing troubles with better players on a team. It is a question of credibility of the team as a society and firm, which has the eventual goal of making efficient choices and avoiding useless costs.

Quantitative analysis of sport performance therefore comes from needs. This needs have developed through the years, along with science and the introduction of new instruments. A great amount of literature has been developed in the last years about these topics, in Europe and especially in the United States. An increasing number of sports are currently being studied, and a great amount of statistics are collected and elaborated in order to get effective results, but the focus is still higher in sports like basketball, baseball, football and hockey, whose American leagues are worldwide recognized and followed. Leagues as the NBA (National Bas-
ket Association), the NFL (National Football Association) or the NBL (National Baseball League) are systems that move a huge amount of money every year (even every game day) and so the interest and expectations of people are very high. This is one of the main reasons why these sports, and especially these leagues, are more covered in terms of statistic studies. But they are not the only ones. Nowadays, there are small firms composed by statisticians that get paid to extract statistics from video recording of a sport event, such as a minor league event of basketball or lacrosse. Thus, everyone playing, or coaching, in any team can have his statistics recorded, and the demand is high, proving that there is an increasing interest over the subject of performance analysis.

As Kubatko et al. in [2], “one of the great strengths of the quantitative analysis of sports is that it is a melting pot for ideas”. It is a field where several academic disciplines and ideas of practitioners and hobbyist from outside academia meet (sometimes efficiently, sometimes not), and where good ideas are implemented immediately into practice. Even an apparently easy task as shooting a goal\(^1\) carries a meaningful amount of importance within the NBA world.

For example, Piette et al. in [3] studied the shooting abilities of NBA players. Their research was focused on evaluating their offensive abilities, given that they try to score with different distances from the basket. One of the most important measures of player’s ability to shoot is field goal percentage (FG%), but it is biased in the sense that does not take count of where the shot starts from. Such a measure can be misleading, since players that play the center position (i.e. closer to opponent’s basket) tend to have higher FG%. The authors thus outlined two new measures of offensive ability that not only account for the distance of the shot from the basket, but can also be used to optimize the length of time each player should spend on court and to decide who amongst those on the court should be given the ball to shoot in certain situations.

Shortridge et al. in [4] characterized and visualized relative spatial shooting effectiveness of basketball players by developing metrics to assess spatial variability in shooting. Assuming that every player takes a unique spatial array of shots, they took advantage of new technologies that made analysis of coordinates possible. They compared and contrasted shooters using a spatial perspective that accounts for the individual shot constellation and the relative performance of all shooters from those spatial locations.

\(^1\)In basketball, goals can be 1, 2 or 3-pointers. Free throws (after fouls) are 1-pointers, while field goals can be 2 or 3-pointers. If a player makes a field goal from within the three-point line, he scores two points. Scoring beyond the three-point line would give three points.
Another effective example of how quantitative analysis has been used in basketball is given by Winston in [5]: “lets assume we have the ball and we trail by two points with little time remaining in the game. Should our primary goal be to attempt a game-tying two-pointer or to go for a buzzer-beating three-pointer to win the game?” (It is reported that this question has been made even in Microsoft job interviews). Using the rule of independent events probabilities, and computing the average probability of scoring a two-pointer, scoring a three-pointer and winning at the overtime\(^2\), Winston derived that it is more efficient (i.e. better chance to winning the game) to go for the three, while most coaches think it is the riskier strategy, and tend to opt for the two\(^3\). Lots of studies and researches have concentrated on basketball, and especially NBA American league, given its great availability of statistics. NBA games have been recorded for several years, and nowadays there are stats for every team and even every player that ever joined the league.

The object of these research is double. First of all, it aims to put order in the field of performance indicators; through the years, a vast body of literature has been created within the sport of basketball, and various method have been developed in order to evaluate performance and attributes of teams and players. Both these two macro-areas (team evaluation and player evaluation) are well covered by numerous methods and algorithms for their assessment, but all of them have their weaknesses (as well as strengths) and statistical question marks. This research aims to present an overview of these indicators and underline these question marks. Second, and core, objective of the research is to understand which are the main aspects a team should concentrate on in order to win its championship (it will be demonstrated later that the focus should be on the “long” period of the league and not on the single game). The method is based on Dean Oliver’s well-recognized work on the factors that determine a teams strengths (and weaknesses) in [1], while the approach through which these factors are studied is the Rough Set Theory, introduced by Pawlak in [12]. This theory is particularly useful when the data are not convenient for traditional, ordinary statistical methods, or, as in this case, when a methodology needs to be proved and improved. As in [11], the rough set approach can lead to identify which, among all the attributes considered to describe a given observation, can be considered as essential in order to state some decision rules; these rules contain the core of the knowledge about some information.

This approach is innovative in the sense that this research represents a first attempt to introduce Rough Set Theory within the field of

\(^2\)Ties do not exist in basketball. If the score is tied at the end of the regular game, teams play multiple overtime periods.

\(^3\)The probability of scoring a three-pointer is greater than the conditional probability of winning in the overtime after scoring a two-pointer.
sport analysis. Rough set-based data analysis methods have already been successfully applied in bioinformatics, economics and finance, medicine, multimedia, web and text mining, software engineering and robotics. So this research makes a step forward in the literature, given the power of this theory to provide certain rules (“If... then...”) that fall from a causality background.

Starting from the definition of team offensive and defensive efficiencies, introduced by Oliver in [1], these quantities are computed for each Italian team that played in Legabasket (the Italian major league of basketball) during the seasons from 2009/10 to 2013/2014. The obtained results are used to create a database, which will be the starting point for the implementation of Rough Set Theory.

The thesis is so implemented: section 2 represents a literature review over the topics covered in the research, in section 3 it is presented the methodology through which the model is explicated. Section 4 presents the results, while section 5 the conclusions.
2 Literature Review

During the last thirty years, basketball at professional level has been deeply analyzed by researchers who wanted to go beyond statistics trying to understand more of what clearly is not just a simple game. Quantitative analysis and all its features have been implemented in order to better comprehend how teams and players behave, so that it could be possible to mathematically control some of the characteristics of the latters and maximize the output of the formers.

This process, as it has been studied, can be developed through various methods, which refer to the specific of the research. That is, approaches have been implemented in order to analyze not only performance on court of teams and players in different time frames, but also, for example, how much a certain player should be paid, or how many minutes he should stay on the court, or even how much time a coach should be given to lead a team to a certain level.

Since the field is vast and big is the body of literature developed, this research would focus its attention on performance analysis. This section represents a literature review on how basketball efficiency, both of teams and players, has been studied through the years, and how the instruments to evaluate it have developed. The interesting fact about these studies is that there are a few shared points of view on what should be taken into account and how it should be weighted, therefore different paths have been taken by the authors. As Berri in [10], when thinking about economics, disagreements about the “truth” abound. “Empirical questions are always given different answers, depending upon the school of thought to which the analyst belongs”. Macroeconomics is complex, and it should not be surprising to see different questions answered in different ways. But in sport world, where data are qualitatively better and most of the times accessible, disagreements should be less frequent. Nevertheless, this does not happen.

One of the first contributions and attempt to evaluate a team’s strength is given by Huang et al. in [8], in 1979. The authors start with the assumption that the production frontier of a basketball team is not conceptually different from that of a firm producing goods: in fact, it stands for the maximum amount of output attainable for a given combination of inputs. Let $F(x)$ be the maximum possible output (the production frontier) and $x$ the given number of inputs. Observed output $Y$ differs from the frontier by a factor $u$:

$$u = \frac{Y}{F(x)}$$

where $u$ is exactly the efficiency through which the outcome is reached,
and $Y$, the observed output, is the ratio of the finals scores of two teams. $u$ is between 0 and 1: in fact, when $u = 1$ the output is equal to the maximum possible so there is maximum efficiency, and when $u = 0$ there is no output at all, therefore null efficiency.

The authors tested their theory using a Cobb-Douglas production function and data from individual games during the 1976-1977 NBA season, with the production frontier as:

$$F(x) = A \times \left( \prod_{i=1}^{8} x_i^{\alpha_i} \right) \times e^{\alpha_9 x_9 + \alpha_{10} x_{10}}$$

where $x_i, i = 1, \ldots, 10$, is a set of statistics like field goal percentages, rebounds and assists, taken as ratios (home team stat./opponent team stat.), but nothing is said about $A$ and $\alpha_j$. Despite the originality of the theoretical background, the results they reached were not consistent, since they found extremely high values of efficiency for the teams they analyzed: the lowest level of efficiency was 0.99849 associated with the New York Knicks, and some teams even showed values greater than 1. Authors provided an explanation that is not really exhaustive, reporting that NBA is a highly competitive market, and therefore coaches and players that do not perform well are quickly ruled out of the market (just as firms that do not produce efficiently would eventually exit the market in a long period perspective). Although the model has some interesting features, results appear not plausible, and there seem to be more than one flaw in this approach, probably due to an incomplete choice of the parameters or a degree of uncertainty that was not taken into account.

In more recent years, authors have concentrated on looking at the problem from a different perspective. “Good” performance metrics should be both correlated with outcomes and consistent over time. As explained by Bradbury in [14]:

One method researchers can use for separating skill from luck is to look at repeat performance of players. If performance is a product of skill, then the athlete in question ought to be able to replicate that skill. If other factors, such as random chance or teammate spillovers are responsible for the performance, then we ought not observe players performing consistently in these areas over time.

The best way to address these issues and develop a consistent indicator for teams’ efficiency is to consider the link between team winning percentage and the specific performance measure, as illustrated in equation (1).

$$Team \ Winning \ Percentage = \alpha_0 + \alpha_1 \times Performance \ Indicator + \epsilon \ (1)$$

where a good performance indicator should exhibit a strong consistency related to equation (1), and therefore should be both correlated
with outcomes and consistent over time.

This regression represents a good test for every performance indicator. As previously outlined, different schools of thought have developed during the years, carrying different ways to evaluate a player or a team. Hereafter these approaches are analyzed. The object of this research is to focus on indicators that evaluate the performance of a team, but frequently there is a fine line between performance indicator for a player and one for a team so it would be inefficient to describe one without referring to the other.

2.1 The Scoring Models

Let’s first turn to a set of models that will be classified as “scoring models”. These kind of models dominate player evaluation in the NBA and, in general, in professional basketball, and such metrics are usually employed when looking at players salaries, the coaches voting for the All-Rookie team, the NBA draft and the allocation of minutes played. Also referred as Linear Weights, these formulas represent a simple way to evaluate a player, multiplying each box score statistic by a weight and equate the weighted sum of player statistics as a measure of the players ability. The easiest approach is the NBA Efficiency index. As indicated in equation (2), this measure adds together the positive actions of a player and subtracts the negatives.

\[
\text{NBA Efficiency} = \text{PTS} + \text{ORB} + \text{DRB} + \text{STL} + \text{BLK} + \text{AST} - \text{TO} - \text{MSFG} - \text{MSFT}
\]

(2)

where:

\[
\begin{align*}
\text{PTS} &= \text{Points scored} \\
\text{ORB} &= \text{Offensive rebounds} \\
\text{DRB} &= \text{Defensive rebounds} \\
\text{STL} &= \text{Steals} \\
\text{BLK} &= \text{Blocks} \\
\text{AST} &= \text{Assists} \\
\text{TO} &= \text{Turnovers} \\
\text{MSFG} &= \text{Missed field goals} \\
\text{MSFT} &= \text{Missed free throws}
\end{align*}
\]

This system essentially states that all good statistics are worth +1 and all bad statistics are worth −1, which in one hand makes the indicator
simple to calculate, but in the other hand makes it way to simplistic.
As in [10], estimating equation (1) with a team’s NBA Efficiency reveals
that this model explains little of a team’s winning percentage\(^4\).

A more complex approach, known as Player Efficiency Rating (PER),
was developed by John Hollinger in 2004. PER is a kind of all-in-one rat-
ing, which attempts to boil down into one number all of a player’s contrib-
ution, measuring a player’s per-minute performance while adjusting for
pace and league average. As the other linear weights, PER formula adds
positive stats and subtract negative ones, all adjusted to a per-minute basis
so that, for example, substitutes can be compared to starters. The
calculation is detailed in equation (3), as implemented in [10]:

\[
\text{PER Player Efficiency Rating} = \\
\text{League Pace} \times \left( \frac{15}{\text{League Average}} \right) \times \left( \frac{1}{\text{Minutes Played}} \right) \times \\
\times [3 \cdot \text{FGM} + \text{AST} \times 0.67 + \left( \text{FGM} \times \left[ 2 - \left( \frac{\text{Team AST}}{\text{Team FGM}} \right) \times 0.588 \right] \right) + \\
+ (\text{FTM} \times 0.5 \times [1 + \left( \frac{\text{Team AST}}{\text{Team FGM}} \right)]) + (\frac{\text{Team AST}}{\text{Team FGM}}) \times \\
\times 0.67]) - (\text{VOP} \times \text{TO}) - (\text{MSFG} \times \text{VOP} \times \text{League DRB%})- \\
- [\text{MSFG} \times \text{VOP} \times 0.44 \times [0.44 + (0.56 \times \text{League DRB%})]] + \\
+ [\text{DRB} \times \text{VOP} \times (1 - \text{League DRB%})] + (\text{ORB} \times \text{VOP} \times \text{League DRB%})+ \\
+ (\text{STL} \times \text{VOP}) + (\text{BLK} \times \text{VOP} \times \text{League DRB%}) - [\text{PF} \times \\
\times [\text{league FTM per PF} - (\text{league FTA per PF} \times 0.44 \times \text{VOP})]]
\] (3)

where:

\[
\text{Pace} = \left[ (\text{Offensive Possessions} + \text{Defensive Possessions}) \times 48 \right] / \\
(\text{Minutes Played}/2)
\]

\[
\text{Possession} = \text{FGM} + 0.44 \times \text{FTM} + \text{TO} - \text{ORB}
\]

\[
\text{FGM} = \text{Field Goals Made}
\]

\[
\text{FTM} = \text{Free Throws Made}
\]

\[
\text{League Average} = \text{Average PER Value}
\]

\[
3\text{FGM} = \text{Three point field goal made}
\]

\[
\text{VOP} = \text{Average points scored per possession for the League}
\]

\[
\text{League DRB%} = \text{Average defensive rebounds divided by average total rebounds}
\]

Player Efficiency Rating is a performance indicator that has its strengths
and weaknesses. As pointed out by Berri in [10] and by Winston in [5],
a player how scores more than 30.4\% on two-point field goals and more

\(^4\)For details, see [10]
than 21.4% on three-point field goals can actually increase his rating by taking more shots. The issue is that these are thresholds that even the worst shooter in the league could overcome. The same problem is related to Hollinger’s Game Score measure, another scoring indicator:

\[
\text{Game Score} = \text{PTS} + 0.4 \times \text{FGM} - 0.7 \times \text{FGA} - 0.4 \times \text{MSFT} + 0.7 \times \text{ORB} + 0.3 \times \text{DRB} + 0.7 \times \text{AST} + 0.7 \times \text{BLK} - 0.4 \times \text{PF} - \text{TO}
\]  

As Winston in [5], a player shooting over 29.2% on two-pointers and over 20.4% on three-pointers will increase his Game Score by taking more shots, and the same happens with NBA Efficiency. Moreover, Berri and Bradbury in [15] proved a correlation of 0.99 between a player’s PER and his Game Score, meaning that they would almost always rank players in the same order, as well as a 0.99 correlation between NBA Efficiency and Game Score. Scoring models, although different in appearance, have the same implications: as long as he exceeds a minimum threshold, the more shots a player takes the better will look his ratings. What these measures fail to estimate is inefficient scoring, which does not help a team win games. Having inefficient scorers take more and more shots does not improve a team’s winning percentage, hence this might be the reason why these indicators, as proved by Berri, are not highly correlated with team outcomes.

Despite the flaws, these measures have their strengths. First of all, being linear equations, they are relatively easy to calculate; secondly, they are widely used, especially PER, which due to its nature can be easily compared between players of different teams and seasons. Specific problems could appear when comparing different generations, that imply different rules and statistical data (for example, the three-point shot entered the league in 1979/80 season), but in general it is possible and it does have a certain appeal on basketball fans and enthusiasts.

Nevertheless, what ultimately wins championships and leagues are not players, but teams. This is the main reason why if the outcome a researcher wishes to understand is the perception of player performance (and therefore his salary or draft position), one of the scoring models may be the preferred choice. But if he wishes to examine wins, these models appear not to fit well.

### 2.2 The Plus-Minus Model

“Plus-minus” (or simply “±”) is a metric that looks at how teams perform with a certain player on the court, how they perform with a certain player off the court, and calculates the overall impact that the player has on team’s outcome. Plus-minus began as a hockey metric, and since
1968 it is an official statistic for NHL. The concept has been broughth in the NBA around 2003, and a few years later the franchise made it a part of its official boxscore statistics. Essentially, plus-minus is computed subtracting all the points the team allows when a certain player is on court from all the points the team scores when he is on court: a player’s plus-minus is how many points better (or worse) the team has done with that man on the court. Intuitively, the more scoring there is in the game, the more effective this statistic becomes, and this is why it is more meaningful in basketball then in hockey.

Plus-minus has earned a good credit during the years for its capability of encompass everything a player does on the court, even things that leagues do not keep official statistics on. Despite that, this measure has some critical problems, converging on the fact that, being basketball a game of five-on-five, the quality of a player’s teammates will impact his plus-minus measure. Basically, a player’s plus-minus is a function of the following factors: player’s ability, ability of his teammates, ability of the teammates who enter the court when he is off, and quality of the opponents. Consequently, this measure does not help a team assign responsibility.

To overcome these problems, researchers have turned to a measure known as “Adjusted plus-minus” (or “Adj±”). The computation involves a regression designed to control for the impact of a player’s teammates and opponents, which solves the fundamental problem of plus-minus, but other issues emerge, regarding specifically statistical significance and consistency. As Berri in [10] and Berri and Bradbury in [15], only 10.2% of the 666 observed player exhibit an Adj± coefficient that is at least twice as large as the corresponding standard error, meaning that most players have an impact on outcomes that cannot be measured significantly, even when more data were involved. Moreover, coefficients appear to be inconsistent over time. Berri and Bradbury in [15] report that only 7% of a player’s adjusted plus-minus can be explained by the same value referred to the previous season, and correlation between these two values is 0.26 (even lower values were found for players who switched teams).

In 2002 Wayne Winston and Jeff Sagarin, both MIT graduates, jointly developed a statistical method to analyze a player’s impact on his team’s ability to produce points. Based on a plus-minus approach, the system works by using the scores of games during periods of time that players are on the floor. Its unique characteristic is that it measures individual and team performance after every lineup change during the game, which is possible only using play-by-play data. The first critique to this approach is presented by Oliver in [1]: plus-minus looks at how teams do with and without a player on the court, essentially breaking down an entire game into segments. In this case, segments represents minutes during which a
given lineup plays. As Oliver notes, teams does not compete to win segments, but games, therefore it is easy to have parts of games in which the point-spread change during those periods does not reflect true levels of competition. Moreover, Oliver focuses his second critique on the interpretation of results: a team seems to do better with a certain player on the court does not mean a lot if you cannot explain how the player is helping.

The greatest contribution by Oliver in [1] is probably the simplest: *there is no Holy Grail of player statistics*. All individual statistics are highly influenced by the context in which they are generated: a player’s teammates or coach make him better or worse, as well as a player’s environment can easily change that player’s performance. Where “environment” is made up of hundreds of variables that cannot be recorded. Despite that, numbers and box scores can tell us a lot about how teams behave and which aspects they should control in order to get good outcomes. Eventually, it is the team that wins and loses, not individuals.

### 2.3 The Win Models

Essentially, what these models try to do is to give consistency to equation (1), that is to try to explain how team reach their outcomes, both good and bad ones, using box score statistics. The two main contributions, given by Berri in [10] and Oliver in [1], are hereafter summarized.

**Berri’s Wins Produced**  This approach is based on Gerrard’s work on soccer in [16]. Berri adapted it to basketball employing three steps to derive the model.

**Step one** Wins are function of points scored and points surrendered.

**Step two** Points scored (or surrendered) are function of conversion rate (or opponents’ conversion rate), i.e. a team’s shooting percentage.

**Step three** Field goal attempts are function of a team’s ability to acquire the ball and what a team does upon gaining possession.

Berri then condensed the aforementioned steps into equation (5):

\[
\text{Winning percentage} = \beta_1 + \beta_2 \times \frac{PTS}{PE} - \beta_3 \times \frac{Opp.PTS}{PA} + \epsilon \quad (5)
\]

where \(PTS\) and \(Opp.PTS\) are points made by the team and its opponents, respectively, and \(PE\) and \(PA\) are possessions employed and possessions acquired\(^5\). Equation (5) states that wins are given by offensive

\(^5\)\(PE = FGA + 0.45 \times FTA + TO - ORB\) and \(PA = Opp.TO + DRB + TMRB + Opp.FGM + 0.45 \times Opp.FTM\)
efficiency and defensive efficiency, where the former is defined as points scored per possession employed and the latter is defined by points surrendered per possession acquired. These definitions of efficiencies were first introduced by Kubatko et al. in [2]; they reflect both how efficient the team is and the pace at which it plays. Given that, in any game, the number of possessions is dictated by both participants and is approximately equal, efficiency with the ball is what ultimately wins.

Berri in [10] has found an adjusted R-Squared = 0.95 for equation (5), regressed over 621 observations. Despite the consistency, the approach seems to raise a fundamental lack. There is strong causality between wins and the factor that imply them, i.e. scoring a lot of points. A more complete and efficient model would try to explain how to get to the step of scoring a lot of points. This is what the next model tries to do.

Oliver’s Factors  Introduced by Dean Oliver in [1], this model aims to figure out which are the most important aspects a team should control in order to win games. These factors can be used to analyze a team’s performance and to better understand a team’s strengths and weaknesses.

1. **Effective field goal percentage.** It is measured by the ratio between two- or three-point field goals and field goals attempted. It gives 50% more credit for making a three-pointer because a three-pointer is worth 50% more points than a two-point field goal.

   \[
eFG_t\% = \frac{FGM_t + 0.5 \times 3FGM_t}{FGA_t} = f1off
   \]

2. **Turnovers per possession.** Turnovers committed per possession.

   \[
   TPP_t = \frac{TOV_t}{POSS_t} = \frac{TOV_t}{FGA_t - OR_t + 0.44 \times FTA_t + TOV_t} = f2off
   \]

3. **Offensive rebounds percentage.** Percentage of rebounds a team gets of their missed shots.

   \[
   OREB_t\% = \frac{OREB_t}{OREB_t + DREB_o} = f3off
   \]

4. **Free throw rate and capacity to get to the foul line.** Free throw rate is measured by foul shots made divided by field goals attempted. It reflects simultaneously the team’s ability to get to the foul line and the ability to make foul shots. This term could be divided into
two terms, one representing how often a team goes to the foul line (relative to shooting from the field) and the other representing how well a team shoots from the foul line.

\[
FTR_t = \frac{FTA_t}{FGA_t} \times \frac{FTM_t}{FTA_t} = \frac{FTM_t}{FGA_t} = f_{4\text{off}}
\]  

(9)

These factors refer to offense, but they can be also calculated for defense (this is why sometimes they refer to them as eight factors). The factor for defense are the following:

5. **Opponent’s effective field goal percentage.**

\[
eFG_o\% = \frac{FGM_o + 0.5 \times 3FGM_o}{FGA_o} = f_{1\text{def}}
\]

(10)

6. **Defensive turnovers forced per possession.**

\[
OTPP_o = \frac{TOV_o}{POSS_o} = \frac{TOV_o}{FGA_o - OR_o + 0.44 \times FTA_o + TOV_o} = f_{2\text{def}}
\]

(11)

7. **Defensive rebounds percentage.**

\[
DREB_t\% = \frac{DREB_t}{DREB_t + OREB_o} = f_{3\text{def}}
\]

(12)

8. **Opponent’s free throw rate and capacity to get to the foul line.**

\[
OFTR_o = \frac{FTM_o}{FGA_o} = f_{4\text{def}}
\]

(13)

As showed, these factors can be calculated both for offense and for defense and provide an effective breakdown of offensive and defensive ratings. Offensively, a team wants to minimize turnovers per possession and maximize the others, while defensively a team aims to maximize turnovers committed and defensive rebounds percentage and minimize the others.

As stated by Oliver in [1], “there really is nothing else in the game”. As reported by Winston in [5], not only these factors are virtually uncorrelated, but running a regression using data from 2007/08 NBA season over these factors can predict the number of wins explaining 91% of the variation of the outcome\(^6\). Similar results were found by Berri\(^7\) in [10],

\(^6\)Winston used the number of games won as dependent variable and the differences between the offensive factors and the respectively defensive ones as dependent variables.

\(^7\)Using 621 observations from a sample concerning NBA seasons from 1987/88 to 2008/09, he computed \textit{Adjusted R-Squared} = 0.936 regressing winning percentage over the eight factors.
where he defines the approach “enlightening”. The only problem with this method is that it does not help measuring the performance of the individual player, since half of the independent variables come from the opponent team and it is not possible to know how each individual employed by a team impacted the opponent’s accumulation of these statistics.

This approach probably represents the most effective way to analyze a team’s performance using box score statistics. Scoring models seem to explain well how a player behave on court, but not how a team wins its games. Plus-minus model teaches the same lesson, and if one is interested in explaining how teams win, definitely Win Models appear to be more effective. Indeed, Oliver’s approach is widely recognized as one of the most productive in the literature of basketball quantitative analysis.

Beside the models already presented, some other important contributions were made by different authors. For example, Martinez in [7] has developed an approach from the starting point that the determinants of the result of a game are the non-scoring variables and the intangibles (leadership of players, sacrifice, team chemistry, etc.) because the result of a game is the difference between the points made by two teams. Then, a component of randomness must be added since it plays an important role in sport. So what he did is introducing a new method to evaluate basketball players using box score statistics, ruling out points made by the player. His model has a quite good consistency, proving that points are not the only tool to evaluate a player’s performance.

A different approach has been developed by Casals and Martinez in [17]. They identified and recorded all variables which may affect the performance of a basketball player from one game to another, then they studied the significance of these variables through mixed models, showing the conditional effects on performance. Their work represents the first attempt in basketball literature to try to understand variation in player performance in a game by game context using this method. Unhappily, interesting results they found can only be applied to players that fit specific characteristics, i.e. a list of only 27 players, obtained through a filtering process, from 2007/08 NBA season.

Another important contribution in the field was presented by Piette et al. in [6] during the MIT Sloan Sports Analytics Conference in 2011. Authors want to tackle the classic problem rising when evaluating player performance, i.e. his interaction with teammates, and study how important was a player relative to all other players, given the five-man units of which a player was member. Basically, they constructed a social network of individual players, where nodes are players and two players are connected if they were a member of the same five-man unit at least once.
The edges are weighted to reflect interdependency between players with respect to their units’ performance, computed using offensive and defensive ratings (points scored/allowed per possession). Then, using random walk, centrality score of each player is computed, therefore determining how central/important is every player relative to all the others. This interesting approach, that involves Bayesian hierarchical models and social network analysis, has a unique issue, that of requiring play-by-play data, i.e. data recording, at least, every single point and substitution made during the game. And these kind of data hardly are available for franchises different from NBA.
3 The Model

The model described in this chapter aims to find and explain which are the determinants that better illustrate why and how a basketball team wins. As specified in the previous section, literature has shown that Oliver’s approach seems to be the most consistent among the ones developed through the years in terms of efficiency of basketball teams: in his contribution he has developed the way to compute eight factors (four for offense and four for defense, as seen before) which are determinant for a team in order to be successful. These factors define a team’s offensive and defensive efficiency. As Oliver in [1], “striving to control those factors leads to a more successful team”. As described in the previous section, these factors have a relative importance with respect to each other; for instance, in order of importance: shooting, turnovers, rebounding and free throws\(^8\). However, nothing is said about the factors related to defense, even if it is rational to presume the same importance and weights.

That said, what is not clear is how these weights should be interpreted; they sum up to 1, and clearly they are not coefficients. Seen from a different point of view, the aim of the model presented here is also to test the importance of these factors. Oliver’s eight factors will be computed for Italian teams in specific seasons of Legabasket. In this way, a database is developed and it includes eight Oliver’s factors (as original developed by the author, four factors for offense and four for defense) and two other factors (synergy and an overall index, both described later) for every team in every season. Despite the clarity and the order, this database is still a bunch of numbers which can be misleading if not uniquely interpreted, especially if each factor is not given the right value. In this field Rough Set Theory finds its importance, and it will be used in order to find implications and sense of the involved factors.

To develop this research, five regular seasons (from 2009/2010 to 2013/2014) are taken into account. Post-seasons (i.e. playoffs) are intentionally omitted due to the so called Theory of Big Dance. Introduced by Vrooman in [9], this theory states that “new science can be effective over regular seasons because the number of games reduces uncertainty over the expected result, but it does not always translate to the post-season”. This is why quantitative analysis is less consistent in the post-season, since “outcomes inherit the randomness of shorter series and knockout tournaments”.

3.1 The database

As said, five seasons of Legabasket are analyzed, more precisely from 2009/2010 to 2013/2014. Data were collected from the official Legabasket-

\(^8\)Their relative weights are, respectively, 0.4, 0.25, 0.20 and 0.15
ket website\textsuperscript{9}, and were later analyzed and handled with Stata 12 software.

For every team which was taking part of each season, the following is recorded or computed:

- \textit{season} - the season which data collected refer to.
- \textit{team} - name of the team. Since it depends on sponsors, which changes over years, the name of the city will be used.
- \textit{wins} - percentage of wins of that team, i.e. games won/games played.
- \textit{Oliver’s eight factors} - effective field goal percentage (and opponents’), offensive and defensive turnovers per possession, offensive and defensive rebounds percentage and free throw rate (and opponents’), i.e. respectively: $f_{1\text{off}}, f_{1\text{def}}, f_{2\text{off}}, f_{2\text{def}}, f_{3\text{off}}, f_{3\text{def}}, f_{4\text{off}}$ and $f_{4\text{def}}$.
- \textit{Synergy} - percentage of players\textsuperscript{10} that have not changed team since previous year.
- \textit{Overall} - index that describes whether a team plays relatively more offensively or defensively.

### 3.2 Oliver’s eight factors

The computational method to find out these variables has been described in the previous section. What is worth underlying is their consistence, presenting some of their important features. First of all, these factors can be computed using box score statistics, which are available for Legabasket. Secondly, there is little (or not at all) correlation between them, as described in Table 1, where each factor is computed as in equations from (6) to (13).

<table>
<thead>
<tr>
<th></th>
<th>$f_{1\text{off}}$</th>
<th>$f_{1\text{def}}$</th>
<th>$f_{2\text{off}}$</th>
<th>$f_{2\text{def}}$</th>
<th>$f_{3\text{off}}$</th>
<th>$f_{3\text{def}}$</th>
<th>$f_{4\text{off}}$</th>
<th>$f_{4\text{def}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{1\text{off}}$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{1\text{def}}$</td>
<td>-0.2673</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{2\text{off}}$</td>
<td>0.0931</td>
<td>-0.2025</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{2\text{def}}$</td>
<td>0.2587</td>
<td>-0.0936</td>
<td>0.3215</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{3\text{off}}$</td>
<td>-0.0724</td>
<td>0.1838</td>
<td>0.0233</td>
<td>0.0484</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{3\text{def}}$</td>
<td>-0.1662</td>
<td>-0.0753</td>
<td>0.0248</td>
<td>-0.2778</td>
<td>-0.0560</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_{4\text{off}}$</td>
<td>0.1162</td>
<td>-0.0850</td>
<td>0.4947</td>
<td>0.4289</td>
<td>0.0654</td>
<td>-0.1570</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>$f_{4\text{def}}$</td>
<td>0.0446</td>
<td>-0.0132</td>
<td>0.3317</td>
<td>0.5419</td>
<td>0.0895</td>
<td>-0.1136</td>
<td>0.2267</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 1: Correlation matrix for Oliver’s factors

\textsuperscript{9}www.legabasket.it

\textsuperscript{10}Only the ten players who have played more minutes overall are included in the computation.
Moreover, these factors, measured for both offence and defense, explain almost 82% of wins percentage of a team, as reported in Table 2. In particular, \( f_{1\text{off}} \) is effective field goal percentage, \( f_{1\text{def}} \) is opponents’ field goal percentage, \( f_{2\text{off}} \) is turnovers allowed, \( f_{2\text{def}} \) is turnovers forced, \( f_{3\text{off}} \) is offensive rebounds percentage, \( f_{3\text{def}} \) is defensive rebounds percentage, \( f_{4\text{off}} \) is free throw rate and finally \( f_{4\text{def}} \) is opponents’ free throw rate (as said, the last two factors reflect not only the ability to make free throws but also the capacity to get to the foul line).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>(Std. Err.)</th>
<th>t-Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_{1\text{off}} )</td>
<td>2.9581</td>
<td>(.3267)</td>
<td>9.05</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_{1\text{def}} )</td>
<td>-3.3675</td>
<td>(.3116)</td>
<td>-10.81</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_{2\text{off}} )</td>
<td>-1.9135</td>
<td>(.3487)</td>
<td>-5.49</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_{2\text{def}} )</td>
<td>2.1975</td>
<td>(.4677)</td>
<td>4.70</td>
<td>0.000</td>
</tr>
<tr>
<td>( f_{3\text{off}} )</td>
<td>.9575</td>
<td>(.2723)</td>
<td>3.52</td>
<td>0.001</td>
</tr>
<tr>
<td>( f_{3\text{def}} )</td>
<td>1.2457</td>
<td>(.3615)</td>
<td>3.45</td>
<td>0.001</td>
</tr>
<tr>
<td>( f_{4\text{off}} )</td>
<td>1.0157</td>
<td>(.2875)</td>
<td>3.53</td>
<td>0.001</td>
</tr>
<tr>
<td>( f_{4\text{def}} )</td>
<td>-.8819</td>
<td>(.2367)</td>
<td>-3.73</td>
<td>0.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>-.4429</td>
<td>(.4178)</td>
<td>-1.06</td>
<td>0.293</td>
</tr>
</tbody>
</table>

Table 2: Estimation of Oliver’s factors coefficients

In order for a team to be efficient, these factors are to be maximized or minimized. Specifically, the factors to be maximized are the ones with a positive coefficient, like for example \( f_{1\text{off}} \) or \( f_{3\text{off}} \), while the factors to be minimized are the ones with a negative coefficient, like \( f_{1\text{def}} \), \( f_{2\text{off}} \) and \( f_{4\text{def}} \). As the regression suggests, p-values of all the factors (but the Intercept) are below 0.05, implying that all the predictors are meaningful. t-Statistics and coefficients show that \( f_{1\text{off}} \) and \( f_{1\text{def}} \) seems to be the most important factors that determine wins, with the latter slightly more that the former, but still the other ones have a significant importance. \( f_{1\text{def}} \) seems to have more statistical importance than \( f_{1\text{off}} \), as well as \( f_{2\text{def}} \) more than \( f_{2\text{off}} \), but this is not enough to express the hypothesis that the defensive phase is more important than the offensive.

Given that these factors depend on teams’ qualities which are not fixed and change over different seasons, their average change over time too. This is explained in Figure 1, that graphs league averages of each of the eight factors from 2009/10 to 2013/14, along with the averages of the same factors referred to a representative team (Siena), which was one of the few teams to compete over every Legabasket season analyzed. As it can be understood intuitively from Figure 1, Siena did well in all five regular seasons. In chronological order, it ranked 1\textsuperscript{st}, 1\textsuperscript{st}, 1\textsuperscript{st}, 5\textsuperscript{th}
and 2nd, and eventually won the title in post-season every year but the last one. Hence, it is reasonable to expect Siena to be above average in every aspect, providing a good team efficiency that allowed the team to reach good performances. Nevertheless, some unexpected results can be deduced.

For instance, $f_{3off}$ is well below average both in 2010/11 and 2011/12 seasons (actually even in 2012/13 when the team gathered a playoff position), meaning a poor offensive rebound rating. Moreover, $f_{3def}$ is below average in 2009/10 and 2013/14, implying a poor efficiency in getting defensive rebounds. The team was not efficient also in other aspects, like getting to the foul line and making free throws ($f_{4off}$) in 2013/14, but allowing the same moves quite a lot ($f_{4def}$) during 2010/11 and 2012/3 seasons.

Figure 1 shows other interesting features. For example, in 2011/12 season, Siena kept, generally, an offense pace lower than previous years, but still managed to win the regular season (maybe because they improved in rebounding). In addition, some of their seasonal peaks were reached in their “worst” year (2012/13), $f_{1def}$ and $f_{3off}$, meaning that there could be some inconsistencies on what should be done in order to reach certain results.
Figure 1: Siena and league average factors values from 2009/10 to 2013/14
Some more “inconsistencies” can be found also through Figure 2. These graphs provide a quick analysis on how teams perform in each factor offensively and defensively, given their rank position, depicting the general range of values. Data are aggregated over all the five seasons: each point on the graphs represents a team, placed on its coordinates (offensive factor, defensive factor). Each point shows, at its right, the final ranking position of the team it represents\textsuperscript{11}.

![Graphs showing offensive and defensive factors from 2009/10 to 2013/14](image)

**Figure 2: Offensive and defensive factors from 2009/10 to 2013/14**

As expected, for the first factor it is easier to find good teams in the lower right corner, and worst in the upper left one, and the same is for the fourth factor. Talking about the second one, the upper left quadrant is where it is easier to find well-ranked teams, and the lower right quadrant is where less efficient teams are expected to be found. Offensive and defensive rebounds ($f_{3off}$ and $f_{3def}$) are both to be maximized, so more efficient teams would be in the upper right corner, and less efficient ones in the lower left corner.

That said, contrary to what should be expected, there are several examples of well ranked teams (i.e. teams that ranked in a high position during the regular season) with low levels of offensive and/or defensive efficiency. And this is true for every factor. This, along with the issues emerged with the case of Siena, means that having good levels of efficiency in every determinant element of the game is not necessary in order

\textsuperscript{11}Note that, since data are aggregated over five season, there will be five points (teams) that ranked 1\textsuperscript{st}, five that ranked 2\textsuperscript{nd} and so on.
to win the League. Therefore, a question rises: what is really necessary in order to win the League? Here is where the contribution of Pawlack and his Rough Set Theory becomes crucial.

3.3 Synergy and Overall indexes

These two indexes are meant to complete what is not said by box score statistics. A team finds its efficiency and, in general, his soul and strength also when it steps off the basketball court. This means that usually a great team is made up by a great group of athletes, people that knows each other well and are highly motivated to reach a specific goal. Therefore, index Synergy is created, and it reflects the roster stability, that is the number of players that are still in the same team from one season to the following, chosen among the ten players who have played more minutes in that team.

The hypothesis is that the more synergy there is within a team, the more that team is efficient, since players got to know each other, both on-court and off-court, and have had more time to practice together. As the former coach of the Chicago Bulls Phil Jackson explains in Eleven Rings: The Soul of Success, “basketball is a great mystery. [...] You can have the perfect mix of talent and the best system of offense and devise a foolproof defensive strategy [...] but if the players don’t have a sense of oneness as a group, your efforts won’t pay off”. In 1989/90 season (the season when Jackson entered the team as a coach) Chicago Bulls could count on one of the most effective and lethal pair of players of the history of NBA, Michael Jordan and Scottie Pippen, and on a great roster itself, but were not able to overcome the NBA semifinals. But once they found a great synergy, and it happened the very next year, they managed to win three straight championships in a row from 1990 to 1993.

Overall index, instead, represents a measure of how a team plays in general, i.e. if it plays relatively better on offense or on defense. For each team, offensive and defensive ratings are computed as in equations (14) and (15), as described by Kubatko et al. in [2]:

$$ORtg_t = \frac{PTS_t}{POSS_t} \times 100$$  \hspace{1cm} (14)

$$DRtg_t = \frac{PTS_o}{POSS_o} \times 100$$  \hspace{1cm} (15)

\[12\] See [18]
where

\[ ORtg_t = \text{Offensive rating} \]
\[ DRtg_t = \text{Defensive rating} \]
\[ PTS_t = \text{Points scored} \]
\[ PTS_o = \text{Points allowed} \]
\[ POSS = \text{Possessions} \]

Equations (14) and (15) describe points scored and allowed per one hundred possessions\(^{13}\). Then, these measures are compared to league averages: if a team does better in offense than in defense compared to league average, its overall index would be “OFF”, otherwise it would be “DEF”. The following formula details the system.

\[
\text{Overall} = \begin{cases} 
\text{OFF} & \text{if } (ORtg_t - ORtg_{\text{league}}) > (DRtg_{\text{league}} - DRtg_t) \\
\text{DEF} & \text{if } (ORtg_t - ORtg_{\text{league}}) \leq (DRtg_{\text{league}} - DRtg_t)
\end{cases}
\]

This measure helps in understanding what is the general game a team plays during the season. The aim is to comprehend which is the approach that pays off more in terms of results, given that big is the difference between an offensive game and a defensive one, since it reflects also the pace at which that team plays.

### 3.4 Rough Set Theory

Rough Set Theory, developed in 1981 by Zdzislaw Pawlak, a Polish mathematician and computer scientist, is a mathematical tool to deal with vague concepts, useful to handle and classify objects characterized by vague, incomplete or lacking information. This approach seems to be useful for this research, where equal results (ranking position) are are accomplished following different paths (efficiency factors).

Data analysis through Rough Set Theory is implemented using a \textbf{decision table} (a database) made up of observations (\textbf{objects}) and their characteristics (\textbf{attributes}), collected in rows. In the decision table, attributes are divided in two different groups called \textbf{condition} and \textbf{decision} attributes. Each row of a decision table produces a \textbf{decision rule}, that states a unique result if some specific conditions are satisfied. As Pawlak in [11], “if a decision rule uniquely determines decision in terms of conditions - the decision rule is \textbf{certain}. Otherwise the decision rule is \textbf{uncertain}”. Every decision rule come along with two conditional probabilities, called \textbf{certainty} and \textbf{coverage} coefficients. The first one describes the conditional probability that an object belongs to the decision

\(^{13}\)Note that the number of possessions per game is approximately equal for both teams, so eventually \(POSS_t = POSS_o\).
class created by the decision rule, given that it undergoes the conditions of the rule. The latter describes the conditional probability of reasons for a specific decision. Both these coefficients satisfy Bayes’ theorem.

The necessary step to apply this theory is to implement a division of the attributes in equivalence classes. In this research, condition attributes are represented by the eight Oliver’s factors, the synergy index and the overall index. In the other hand, decision attribute is only one, i.e. the ranking position of a team at the end of regular season. Since these attributes are numeric values, it is useful to group them into different classes, in order to avoid an excessive detail of the information set and to prevent an extreme specification of the decision rules. Therefore, values of each attribute are aggregated in different equivalence classes, described hereafter.

Condition attributes:

- Oliver’s eight factors: each value (for each team) of these attributes is grouped in one class. Three classes are created, named “LOW”, “MEDIUM” and “HIGH”. The first one includes all the values of that specific attribute (factor) which are lower than (or equal to) the 33rd percentile of the set of values, considering seasons. The second one includes all the values between the 33rd and the 66th percentile, while the class “HIGH” includes the values greater than (but different to) the 66th percentile of the set. For the factors to be minimized (f1def, f2off and f4def), the contrary applies: greater values of the factor correspond to lower values on the classification.

- Synergy: again, the classes created are three, i.e. “LOW”, “MEDIUM” and “HIGH”. Being an index which take values between 0 and 1, synergy is ranked low when it is less than (or equal to) 0.33, medium when it is between 0.33 and 0.66, and high when it is greater than (but different to) 0.66. Then, a complementary class is added for those teams that are “new” in Legabasket, i.e. teams that just came got promoted from the inferior league\footnote{Rosters for past seasons of Lega2 are not available.}. These teams join the class called “NEW”.

- Overall index: two classes are created, “OFF” and “DEF”, and the values for each team are grouped following the formula described above.

Decision attribute:

- Ranking position: four classes are created. “TITLE” includes the first four positions in the final rank, so positions from the 1st to the 4th. It is called “TITLE” because, even without guaranteeing the win of the title, ending up among one of these positions put a team
in an advantageous position to face the playoffs\textsuperscript{15}, i.e. these teams will have better probabilities to win the title over the other teams in the playoffs. Class “\textit{PLAYOFF}” includes teams ranked from $5^{th}$ to $8^{th}$. These teams still have possibilities to win the title, but being in these positions they cannot count on home court advantage during playoffs. Class “\textit{CENTER}” includes teams ranked from $9^{th}$ to $13^{th}$ position, while the last class, “\textit{RELEGATION}”, includes the last teams, ranked from $14^{th}$ position on\textsuperscript{16}. This last class is called “\textit{RELEGATION}” since teams that belong to it struggle, usually until the very last game, to avoid relegation to inferior league, that happens to the last ranking team.

As it will be explained in the next chapter, there are some inconsistent results that appear from the database before results are found. For example, not all team that ranked in the class “\textit{TITLE}” showed high values in the other conditional attributes. Some factors appear to be weaker than others, meaning that they do not affect so much the final outcome.

Essentially, there seem to be a lot of different paths leading to different results, like there exists possibly infinite roads that lead to a limited number of places. For instance, let’s examine the different skill that the five team that won the regular season from 2009/10 to 2013/14 showed, as in Table 3.

Milano and Varese, that won the last two regular seasons, played basically a defensive basketball, and both teams started the season with a roster where very little players were still there from the previous one. Conversely, Siena, that was able to win three regular seasons in a row from 2009/10 to 2011/12 (and eventually to win also playoffs), used to play a more offensive basketball, and was able to do it effectively taking advantage of a core number of players that represented the team for a few year uninterrupted. The team was good enough to outclass the opponents in almost every field factor, both on offense and in defense, even if showing a relatively poor rebounding capacity throughout the years. On the other hand, Milano and Varese followed almost opposite paths:

\textsuperscript{15}In Legabasket, first eight team ranked in regular season compete for the title through playoffs. In the first phase (quarterfinals), $1^{st}$ ranked faces the $8^{th}$, the $4^{th}$ faces the $5^{th}$, and then there is $2^{nd}$ vs $7^{th}$ and $3^{rd}$ vs $6^{th}$. Second phase (semifinals) sees the winner of the first match against the winner of the second, and the winner of the third against the winner of the last one, and so on through the finals. Quarterfinals are best-of-5, while semifinals and finals are best-of-7. The first four ranked teams in the regular season find themselves in a better position than the others exploiting the home court advantage, since they have the right to play more games in their own field.

\textsuperscript{16}This class have variable number of values since the total number of teams competing in Legabasket has not been fixed over the years. For instance, in 2009/10 season there were 15 teams, 17 in the 2011/12 season and 16 in the others.
Milano kept an amazing rhythm in offence, outscoring the opponents in every offensive factor but still maintaining a defensive approach to the game, while Varese found its balance in the defensive side, allowing very little to the opponents and still keeping an efficient offensive phase.

Table 4 shows, conversely, which were the problems of teams that during these season were relegated to Italian Lega2, the inferior league. Every team shoted very poorly from the field, way below the league average, and every team had a low roster stability or were new in the league. But here the common features end. Montegranaro and Brindisi both played an offensively-oriented basketball, which led them to have a good capacity on getting to the foul line and making free throws, but they had terrible rebounding skills and, as said, a field goal percentage poor on offense and high on defense. Monferrato and Ferrara both showed a more defensive play, doing good in not allowing a high field goal percentage to the opponents. But they lacked significantly in getting (offensive) rebounds and allowing a lot of turnovers. Lastly, Biella, apart from defensive rebounds and field goal percentage allowed (which anyhow reflect its defensive attitude) was very ineffective in every factor, probably also because it was new in the league and lacked of experience.
<table>
<thead>
<tr>
<th>Season</th>
<th>Team</th>
<th>$f_{1\text{off}}$</th>
<th>$f_{1\text{def}}$</th>
<th>$f_{2\text{off}}$</th>
<th>$f_{2\text{def}}$</th>
<th>$f_{3\text{off}}$</th>
<th>$f_{3\text{def}}$</th>
<th>$f_{4\text{off}}$</th>
<th>$f_{4\text{def}}$</th>
<th>Synergy</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/10</td>
<td>Siena</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>OFF</td>
</tr>
<tr>
<td>2010/11</td>
<td>Siena</td>
<td>HIGH</td>
<td>HIGH</td>
<td>LOW</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>OFF</td>
</tr>
<tr>
<td>2011/12</td>
<td>Siena</td>
<td>HIGH</td>
<td>HIGH</td>
<td>LOW</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>OFF</td>
</tr>
<tr>
<td>2012/13</td>
<td>Varese</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>DEF</td>
<td></td>
</tr>
<tr>
<td>2013/14</td>
<td>Milano</td>
<td>HIGH</td>
<td>LOW</td>
<td>HIGH</td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>LOW</td>
<td>LOW</td>
<td>DEF</td>
</tr>
</tbody>
</table>

Table 3: Teams ranked FIRST in regular season

<table>
<thead>
<tr>
<th>Season</th>
<th>Team</th>
<th>$f_{1\text{off}}$</th>
<th>$f_{1\text{def}}$</th>
<th>$f_{2\text{off}}$</th>
<th>$f_{2\text{def}}$</th>
<th>$f_{3\text{off}}$</th>
<th>$f_{3\text{def}}$</th>
<th>$f_{4\text{off}}$</th>
<th>$f_{4\text{def}}$</th>
<th>Synergy</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/10</td>
<td>Ferrara</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>HIGH</td>
<td>LOW</td>
<td>DEF</td>
</tr>
<tr>
<td>2010/11</td>
<td>Brindisi</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
<td>HIGH</td>
<td>HIGH</td>
<td>HIGH</td>
<td>NEW</td>
</tr>
<tr>
<td>2011/12</td>
<td>Monferrato</td>
<td>LOW</td>
<td>HIGH</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>LOW</td>
<td>NEW</td>
<td>DEF</td>
</tr>
<tr>
<td>2012/13</td>
<td>Biella</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>DEF</td>
</tr>
<tr>
<td>2013/14</td>
<td>Montegranaro</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>HIGH</td>
<td>LOW</td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>LOW</td>
<td>LOW</td>
<td>OFF</td>
</tr>
</tbody>
</table>

Table 4: Teams ranked LAST in regular season
As seen, the map seems confusing. Rough Set Theory, as implemented in the next chapter, will be useful to bring clarity and certain answers. But first, a few hypothesis can be raised. As explained, it is rational to expect that teams with higher factors would rank better. Among these factors, it is reasonable to think that teams running for the title would exhibit exceptional shooting efficiencies, and conversely team that try to avoid relegation should have a shooting percentage “less worse” than the others in the same position. The same is valid for the other factors, and it is also rational to expect that team with greater synergy would do better, while nothing certain can be said a priori about the overall index.
4 Results

The results of this research are obtained using ROSE2, a software based on Rough Set method. As previously explained, it helps to obtain decision rules of the form “CONDITION ⇒ DECISION” from the decision table, i.e. the dataset. Moreover, two indexes are computed, coverage and certainty, that reflect the goodness of every single decision rule. Rough Set Theory (or RST) is here implemented with the aim of finding inconsistencies in the database, like the ones that were earlier discussed, and drawing some interesting decision rules.

As computed by ROSE2, several decision rules are found, with different levels of coverage and certainty. Hereafter the outcomes of the research will be showed, represented by the main decision rules that emerged, therefore the most significant ones in terms of certainty and coverage. This section is divided in four parts. In the first one some basic definitions about RST are recalled. In the second one, some generally accepted basketball myths are collected and interpreted depending on what emerges from this study. In the third one it is presented a list of the main rules that have been found using ROSE2, rules with a high degree of certainty in order to be more consistent. The last part is a collection of indirect rules, that are presented as possible strategies for a team in order to achieve a precise ranking position during the regular season.

4.1 Coverage and certainty

As described in the previous section, data analysis through Rough Set Theory is implemented using a decision table, composed of observations (objects) and their characteristics (attributes). Attributes are divided in condition ones and decision ones: decision rules are expressed as a decision attribute which is verified when some conditions are valid. Two indicators, called coverage and certainty, give a measure of the goodness of the rule.

Hereafter some basic definitions in Pawlak’ Rough Set Theory are recalled (for details, see [11] and [12]). Assuming that objects characterized by the same information are indiscernible, a finite nonempty universe $U$ of elements $\{x_1, x_2, \ldots, x_N\}$ is related to the finite set $A$ of $k$ attributes $\{a_1, a_2, \ldots, a_k\}$ by the domain sets $V_{a_i}$ (for each $i = 1, \ldots, k$), i.e. the sets of all the values of each attribute $a_i$. The pair $S = (U, A)$ is called an information system. For each $B \subseteq A$, the universe $U$ is split into a family of equivalences classes (called elementary sets) through the indiscernibility relation $IND(B)$ stating that two objects $x_i$ and $x_j$ in $U$ cannot be distinguished with reference to the set of attributes in $B$. That is, the two objects are considered to be indiscernible or equivalent if and
only if they have the same values for all attributes in the set $B$. When a concept $X$, that is a set $X \subseteq U$, is composed of objects that are all in elementary sets, then all its objects can be distinguished in terms of the available attributes; otherwise, $X$ is roughly defined. Namely, $X$ can be approximated by the sets $B$-lower and $B$-upper approximations of $X$. More precisely, by denoting with $[x]_B$ the equivalence class containing $x$, then the $B$-lower approximation of $X$ is

$$
\underline{B}(X) = \{x \in U : [x]_B \subseteq X\}
$$

that is $\underline{B}(X)$ is composed of all the elementary sets that are included in $X$ in a not ambiguous way, while the $B$-upper approximation of $X$

$$
\overline{B}(X) = \{x \in U : [x]_B \cap X \neq \emptyset\}
$$
is the set of all the elementary sets that have a nonempty intersection with $X$. Therefore, a concept $X$ is called exact with respect to $B$ if the boundary region

$$
BN_B(X) = \underline{B}(X) - \overline{B}(X)
$$
is empty, otherwise is said to be rough with respect to $B$.

Every object is described by condition attributes and decision attributes, therefore the information system will be called a decision table and will be denoted by $S = (U, C, D)$ where $C$ and $D$ are respectively the disjoint sets of condition and decision attributes. The decision rule induced by $x$ in $S$ is, in short, the sequence $C \rightarrow_x D$. With reference to this decision rule, it is possible to set the definitions of certainty factor $cer_x(C, D)$ and coverage factor $cov_x(C, D)$ as it follows

$$
cer_x(C, D) = \frac{|C(x) \cap D(x)|}{|C(x)|}, \quad cov_x(C, D) = \frac{|C(x) \cap D(x)|}{|D(x)|},
$$

respectively, where $C(x)$ denotes the set of condition attributes and $D(x)$ represents the set of decision attributes. These two indexes admit a probabilistic meaning: the certainty factor of the decision rule represents the conditional probability that $y \in D(x)$ conditionally to the assumption that $y \in C(x)$, while the coverage factor represents the conditional probability that $y \in C(x)$ conditionally to the assumption that $y \in D(x)$. In this sense, as stated before, both these coefficients satisfy Bayes’ theorem. It is possible also to study indirect decision rules, i.e. $D \rightarrow_x C$, and it is easy to prove that $cer_x(C, D) = cov_x(D, C)$ and $cov_x(C, D) = cer_x(D, C)$.

### 4.2 (False) Basketball myths

In this section some “basketball myths” will be presented, and Rough Set Theory, along with the dataset created, will help to state, with a
certain level of certainty, whether or not these rules are reliable, or even lacking some crucial aspects. Of course, the discussions are based on the information utilized, for instance those regarding five different seasons of Italian Legabasket. These myths are statements, made by experts in the field of quantitative sport analysis, that are generally assumed as indisputable. Every myth will be followed by a choice of rules generated by ROSE2. The choice was made taking those rules with highest levels of certainty and coverage, and involve the factors that are directly or indirectly mentioned in the myth, so that it is discussed and possibly proved or implemented. Since rules are supported by observations, the list of observations (team) that generated each rule is detailed in Appendix.

Hypothesis 1 - “The team that shoots the best wins the most. Shooting and stopping the opponents’ shooters are the bottom lines.”\footnote{Oliver in [1]}

This statement simply puts the offensive and defensive shooting percentage as the most important factors a team should control in order to win games. In other words, the better a team shoots (and the better it blocks opponents’ attempts), the more it wins. Table 5 offers an objective view on how offensive and defensive shooting percentage, i.e. $f_{1\text{off}}$ and $f_{1\text{def}}$, affect the ranking position of a team. Every rule in Table 5 has certainty = 100.00\% (i.e. the conditional probability that a team $y$ belongs to $D(x)$ given that it belongs to $C(x)$ is equal to 1), so only coverage factor is showed (i.e. the conditional probability that a team $y$ belongs to $C(x)$ given that it belongs to $D(x)$).
Almost all the teams that go for the title show an offensive and defensive shooting factor which is on average or above. Precisely, as in rule 1, if a team has a high shooting percentage \( f_{1\text{off}} = \text{HIGH} \) and keeps its opponents to a low one \( f_{1\text{def}} = \text{LOW} \), then it will end in title class, if it keeps also a high offensive rebounds rate. 25\% of teams that finished the season in title class followed a similar strategy, and conversely every team that followed this strategy ended the season in title class (certainty = 100\%). A team can reach the same result keeping a lower defensive shooting percentage \( f_{1\text{def}} = \text{MEDIUM} \) but it needs to focus strongly on offense \( f_{1\text{off}} = \text{HIGH} \) and \( \text{Overall} = \text{OFF} \) and start with a roster stability which is at least medium (rule 2).

Conversely, if a team keeps a good defense and an average shooting rate \( f_{1\text{def}} = \text{HIGH} \) and \( f_{1\text{off}} = \text{MEDIUM} \) along with an average defensive rebounds rate and a defensive play (rule 3) or a low capacity of avoiding turnovers (rule 4), at most it can aim to a playoff position. Lastly, as in rule 5, it is difficult to avoid relegation for a team if it shoots badly and defend badly as well \( f_{1\text{off}} = \text{LOW} \) and \( f_{1\text{def}} = \text{LOW} \), even more difficult if it has a low roster stability and allows an average number of turnovers.

\[ \Rightarrow \text{Implication 1 - Shooting and blocking opponents' shooters are important factors. If a team wants to fight for the title, the latter is more important than the former, while the opposite is true if a team aims for} \]

<table>
<thead>
<tr>
<th>Rule</th>
<th>IF (Condition attributes)</th>
<th>THEN (Decision attribute)</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( f_{1\text{off}} = \text{HIGH} ) &lt;br&gt;( f_{1\text{def}} = \text{HIGH} ) &lt;br&gt;( f_{3\text{off}} = \text{HIGH} )</td>
<td>Title</td>
<td>25.00%</td>
</tr>
<tr>
<td>2</td>
<td>( f_{1\text{off}} = \text{HIGH} ) &lt;br&gt;( f_{1\text{def}} = \text{MEDIUM} ) &lt;br&gt;( \text{Syn} = \text{MEDIUM} ) &lt;br&gt;( \text{Overall} = \text{OFF} )</td>
<td>Title</td>
<td>15.00%</td>
</tr>
<tr>
<td>3</td>
<td>( f_{1\text{off}} = \text{MEDIUM} ) &lt;br&gt;( f_{1\text{def}} = \text{HIGH} ) &lt;br&gt;( f_{3\text{def}} = \text{MEDIUM} ) &lt;br&gt;( \text{Overall} = \text{DEF} )</td>
<td>Playoff</td>
<td>20.00%</td>
</tr>
<tr>
<td>4</td>
<td>( f_{1\text{off}} = \text{MEDIUM} ) &lt;br&gt;( f_{1\text{def}} = \text{HIGH} ) &lt;br&gt;( f_{2\text{off}} = \text{LOW} )</td>
<td>Playoff</td>
<td>15.00%</td>
</tr>
<tr>
<td>5</td>
<td>( f_{1\text{off}} = \text{LOW} ) &lt;br&gt;( f_{1\text{def}} = \text{LOW} ) &lt;br&gt;( f_{2\text{off}} = \text{MEDIUM} ) &lt;br&gt;( \text{Syn} = \text{LOW} )</td>
<td>Relegation</td>
<td>33.33%</td>
</tr>
</tbody>
</table>

Table 5: How offensive and defensive shooting percentage affect ranking
Hypothesis 2 - “Teamwork is a way of coordinating teammates’ interactions to increase a team’s chances of scoring.”

Although these sentence can be interpreted in different ways, what is central in the idea of increasing the performance is the concept of teamwork. Even if teamwork is a variable which is difficult to quantify within a team, an attempt has been made with the attribute Synergy. Synergy reflects roster stability and teamwork, and the statement implies that teamwork is a good starting point for a team in order to be more successful. In Table 6 it is summarized how the factor synergy affects the ranking position of a team during regular season. Again, certainty factor is omitted since every rule in Table 6 has certainty = 100%.

<table>
<thead>
<tr>
<th>Rule</th>
<th>IF</th>
<th>THEN</th>
<th>Coverage</th>
</tr>
</thead>
</table>
| 6    | Syn = MEDIUM  
      | f1off = HIGH  
      | f4def = HIGH | Title | 25.00% |
| 7    | Syn = MEDIUM  
      | f4off = HIGH  
      | Overall = DEF | Playoff | 15.00% |
| 8    | Syn = LOW    
      | f1def = HIGH  
      | f3def = MEDIUM | Playoff | 20.00% |
| 9    | Syn = MEDIUM  
      | f1def = LOW | Centre | 15.00% |
| 10   | Syn = LOW    
      | f1off = LOW   
      | f3off = HIGH  
      | Overall = OFF | Relegation | 26.67% |

Table 6: How synergy affects ranking

Looking at the rules generated in Table 6, it is possible to notice that teams with an average roster stability (Syn = MEDIUM) find themselves in a better position than those with a low one. Precisely, as in rule 6, a team with a medium synergy can aim for the title class if it keeps a good shooting percentage and does not let the opponents to the foul line a lot (f1off = HIGH and f4def = HIGH). As in rule 7, it can also reach playoffs if it plays defensively along with a good capacity to get to the foul line and make free throws. A team can also start with a low roster

\(^{18}\)Oliver in [1]

\(^{19}\)Note that the dataset reported only two cases (out of 80) of teams with high synergy.
stability, but if it plays very well defensively (Overall = DEF and f1def = HIGH) and keeps an average defensive rebounds rate (rule 8), then it can reach playoffs anyway.

However, synergy is not all. In fact, as in rule 9, a team with a medium roster stability will end up in centre zone if it allows the opponents to make a lot of shots from the field (f1def = LOW). However, as well as rule 7, this rule has a relatively low coverage, since only 15% of teams that were in centre class showed those features. As said, a low roster stability can be balanced by good defensive skills in order to reach playoffs. But if it is together with an overall defensive play and the shooting percentage is poor (Overall = DEF and f1off = LOW), rule 10 sais that relegation class cannot be avoided, even if the team is good on getting offensive rebounds (however, just 26.67% of team that ended in relegation class were showing this characteristics).

⇒ Implication 2 - Synergy, or roster stability, is an important factor that a team should look at if it wants to be successful. Teams in relegation class show low synergy or new rosters, while a team running for the title should, at least, show a medium level of roster stability.

Hypothesis 3 - “Rebounds do have a substantial impact on wins.”

Over the years, the role and weight of rebounding has been subject to a deep and precise analysis. One of the first authors to underline the importance of rebounds was David Berri, already cited in section 2. His Wins Produced method, that has applied to single players, has pointed out the essential role of rebounding in order to win games. Table 7 helps to understand the consistency of offensive and defensive rebounds within Legabasket. Again, certainty is 100.00% for every rule generated by ROSE2 in Table 7.

\[20\]www.wagesofwins.com/faq/
Table 7: How offensive and defensive rebounds affect ranking

As reported in Table 7, rule 11 says that if a team keeps an average rebound rate ($f_{3\text{off}} = \text{MEDIUM}$ and $f_{3\text{def}} = \text{MEDIUM}$) along with an average capacity of blocking opponents’ shots and a good capacity to get to the foul line ($f_{1\text{def}} = \text{MEDIUM}$ and $f_{4\text{def}} = \text{HIGH}$) it will end in title class. The importance of rebounds seems not very big. A team, despite a low offensive rebound rate ($f_{3\text{off}} = \text{LOW}$) and an average defensive rebound rate ($f_{3\text{def}} = \text{MEDIUM}$) can reach the playoffs even with a low roster stability if it plays defensively, as stated by rule 12. Both rules 11 and 12 have a coverage factor equal to 15.00%. Rule 13 shows that it can end in the centre zone even if its rebound rates are high and the number of turnovers committed is on average.

Thus, keeping high rebound rates does not seem so important for a team to win, as well as it does not necessarily imply a good ranking position at the end of the season. This idea is strengthened by rule 14, which states that, even with a good rebounding capacity ($f_{3\text{off}} = \text{MEDIUM}$ and $f_{3\text{def}} = \text{HIGH}$), a team can end in relegation class if its opponents shoot very well ($f_{1\text{def}} = \text{LOW}$). As in rule 15, its position would not change with a lower rebounding percentage, as long as its defense is not effective.

$\Rightarrow$ Implication 3 - Rebounding does not win championships. Even if teams that want to win the league should maintain a good rebounding percentage, this is not compulsory nor essential.
**Hypothesis 4** - A good capacity in ball handling is more important than the ability of getting to the foul line and scoring free throws. The four factors, as described by Oliver and listed in the previous section, appear in order of importance. Now, the second factor \( f_{2 \text{off}} \), i.e. turnovers committed per possession, represents very well the capacity of a team to handle the ball, since the more a team is capable to do it, the lower is the turnover ratio per possession. The fourth factor \( f_{4 \text{off}} \) represents both the capacity of getting to the foul line and the ability to make free throws. In Table 8 results regarding how these factors affect ranking position are listed, so it is possible to compare them.

<table>
<thead>
<tr>
<th>Rule</th>
<th><strong>IF</strong></th>
<th><strong>THEN</strong></th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Condition attributes)</td>
<td>(Decision attribute)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>( f_{2 \text{off}} = \text{HIGH} ) &lt;br&gt;( f_{4 \text{off}} = \text{HIGH} ) &lt;br&gt;( \text{Syn} = \text{MEDIUM} )</td>
<td>Title</td>
<td>15.00%</td>
</tr>
<tr>
<td>17</td>
<td>( f_{2 \text{off}} = \text{MEDIUM} ) &lt;br&gt;( f_{4 \text{off}} = \text{HIGH} ) &lt;br&gt;( \text{Syn} = \text{NEW} )</td>
<td>Playoff</td>
<td>15.00%</td>
</tr>
<tr>
<td>18</td>
<td>( f_{2 \text{off}} = \text{MEDIUM} ) &lt;br&gt;( f_{4 \text{off}} = \text{HIGH} ) &lt;br&gt;( f_{3 \text{def}} = \text{LOW} )</td>
<td>Playoff</td>
<td>15.00%</td>
</tr>
<tr>
<td>19</td>
<td>( f_{2 \text{off}} = \text{MEDIUM} ) &lt;br&gt;( f_{4 \text{off}} = \text{LOW} ) &lt;br&gt;( f_{4 \text{def}} = \text{MEDIUM} )</td>
<td>Centre</td>
<td>12.00%</td>
</tr>
</tbody>
</table>

Table 8: How committing turnovers, reaching the foul line and free throws percentage affect ranking position

As before, all rules in Table 8 have 100% certainty. Rule 16 says that if a team keeps a high level both in ball handling and in the capacity of getting to the foul line, together with an average roster stability, it will end the regular season in title class with **coverage** = 15%, that is 15% of teams that were in title class implemented this strategy. Some more interesting aspects emerge from the following rules: rules 17 and 18 say that committing an average number of turnovers and having a good ability of getting to the foul line \( f_{2 \text{off}} = \text{MEDIUM} \) and \( f_{4 \text{off}} = \text{HIGH} \) would take a team to a playoff position despite an average defensive rebound rate \( f_{3 \text{def}} = \text{MEDIUM} \) or being a newly promoted team. Both these rules have a 15% coverage.

Conversely, rule 19 states that if the number of turnovers committed is still on average but the capacity of reaching the foul line drops \( f_{2 \text{off}} = \text{MEDIUM} \) and \( f_{4 \text{off}} = \text{LOW} \), a team will finish in centre zone, if it is also on average in the ability of letting the opponents far from the

---
21 As the weights of the four factors, reported in the previous section, imply.
foul line \( (f_{4\text{def}} = \text{MEDIUM}) \). Clearly, \( f_{4\text{off}} \) seems more important in affecting the ranking position with respect to \( f_{2\text{off}} \), at least looking at certainty factors; in fact, more is left since these rules have a relatively low coverage.

⇒ Implication 4 - The ability of getting to the foul line and making free throws should be given more importance than that of avoiding turnovers.

Hypothesis 5 - “Defense probably doesn’t win championships.”

There has been an ongoing debate about how a basketball team should interpret its own game, if more offensively or more defensively. As reported in the previous sections, there are plenty of instances of teams that won championships playing a purely offensive basketball, but also there are, conversely, examples of teams that did it playing more defensively. Table 9 can help in solving this issue, with reference to our database (note that all the rules have certainty = 100%)

<table>
<thead>
<tr>
<th>Rule</th>
<th>IF (Condition attributes)</th>
<th>THEN (Decision attribute)</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>( \text{Overall} = \text{OFF} ) &lt;br&gt;( f_{1\text{off}} = \text{HIGH} )&lt;br&gt;( f_{2\text{def}} = \text{HIGH} )&lt;br&gt;( \text{Syn} = \text{MEDIUM} )</td>
<td>Title</td>
<td>20.00%</td>
</tr>
<tr>
<td>21</td>
<td>( \text{Overall} = \text{OFF} ) &lt;br&gt;( f_{2\text{def}} = \text{MEDIUM} )&lt;br&gt;( f_{3\text{def}} = \text{HIGH} )</td>
<td>Centre</td>
<td>24.00%</td>
</tr>
<tr>
<td>22</td>
<td>( \text{Overall} = \text{DEF} ) &lt;br&gt;( f_{1\text{def}} = \text{HIGH} )&lt;br&gt;( f_{3\text{def}} = \text{MEDIUM} )&lt;br&gt;( \text{Syn} = \text{LOW} )</td>
<td>Playoff</td>
<td>20.00%</td>
</tr>
<tr>
<td>23</td>
<td>( \text{Overall} = \text{OFF} ) &lt;br&gt;( f_{1\text{off}} = \text{LOW} )&lt;br&gt;( f_{3\text{off}} = \text{HIGH} )&lt;br&gt;( \text{Syn} = \text{LOW} )</td>
<td>Relegation</td>
<td>26.67%</td>
</tr>
<tr>
<td>24</td>
<td>( \text{Overall} = \text{DEF} ) &lt;br&gt;( f_{1\text{def}} = \text{LOW} )&lt;br&gt;( f_{2\text{def}} = \text{MEDIUM} )</td>
<td>Relegation</td>
<td>26.67%</td>
</tr>
</tbody>
</table>

Table 9: How offensive or defensive play affect ranking position

As reported in Table 9, playing an offensive basketball can be productive or counter-productive depending on the other strategies that are implemented with. For instance, rule 20 states that if an offensive basketball is followed by a good offensive shooting percentage \( (f_{1\text{off}} = \text{HIGH}) \), a good capacity on forcing turnovers \( (f_{2\text{def}} = \text{HIGH}) \) and a medium roster stability \( (\text{Syn} = \text{MEDIUM}) \), then the team will reach the title class.

\(^{22}\) Oliver in [1].
But if the same offensive strategy is followed by different strengths (for instance, an average capacity of forcing turnovers, \( f_{2\text{def}} = \text{MEDIUM} \), and a good defensive rebound rate, \( f_{3\text{def}} = \text{HIGH} \)), then it does reveal counter-productive, leading a team in the centre zone (rule 21).

Conversely, an overall defensive play easily takes a team to playoffs if supported by a good defensive shooting percentage (\( f_{1\text{def}} = \text{HIGH} \)), an average defensive rebound rate (\( f_{3\text{def}} = \text{MEDIUM} \)) and even a low roster stability (rule 22). Rules 23 and 24 show how, in order to be successful, a team should implement strategies which are consistent with their overall playing tendency. As rule 23, a team that plays more offensively but has low shooting percentage (\( f_{1\text{off}} = \text{LOW} \)), high offensive rebound rate (\( f_{3\text{off}} = \text{HIGH} \)) and low synergy will end up in relegation class. As rule 24, a team that plays more defensively but shows poor capacity on blocking opponents’ shots (\( f_{1\text{def}} = \text{LOW} \)) and is on average on forcing turnovers (\( f_{2\text{def}} = \text{MEDIUM} \)), will end the season in relegation class as well. These last two rules have coverage = 26.67%.

⇒ Implication 5 - During the regular season, defense helps getting to the playoff more than offense does. As Vrooman in [9], what probably happens is that “offense sells tickets and defense wins championships”. However, nothing can be said about the number of tickets sold by the offense.

4.3 General rules

Some generally accepted and recognized rule have been discussed and slightly modified. Of course, it should be kept in mind that these results refer to data collected within Legabasket, the italian championship, so they do not necessarily apply to systems and leagues different from this one. Now, let us consider some general rules (implications) that have been found using ROSE2. These rules were chosen among the ones with highest level of certainty and coverage. For a more analytic point of view, see Table 10. Again, see Appendix for the list of observations that contributed to generate each specific rule.

25. A high level of effective field goal percentage on offense, along with a low turnover rate, leads a team to the best position in order to strive for the title.

This rule states that a team that keeps a high field goal percentage (i.e. a high level of field goals made over field goals attempted) and has the capacity of letting few turnovers in regular season, likely will be in a good position to fight for the title in post-season. The rule is 100% certain, which means that all the team that showed these two features ended up in title class.

26. A high level of effective field goal percentage on offense, together with a good capacity of keeping the opponents far from the foul
line and an average defensive field goal percentage, leads a team to aim for the title.

If a team if able to shoot very well from the field, and, by the defensive side, it is able not only to not let the opponents to keep their field goal percentage high but also to make few fouls (or to keep them far from the foul line), then it ends the regular season in title class.

27. A high level of effective field goal percentage on offense, along with a high offensive rebounds percentage and a capacity of getting to the foul line and making free throws at least on average, leads a team to fight for the title.

Along with a high offensive field goals percentage, teams that are able to keep an outstanding offensive rebounds rate and that are able to keep an average level of reaching the foul line and free throws percentage, end the regular season in title class.

28. If a team plays a defensive basketball, keeps the opponents to a low field goal percentage and gets to the foul line a lot, for sure it will participate to the playoffs.

A team that plays better on defense then on offense, that take advantage of its good defense letting the opponents to a low field goal percentage and that is able, in offense, to get a lot of fouls, realizing the free throws, achieves a position for the playoffs.

29. If a team plays a defensive basketball and allows few turnovers will get to the playoffs, even if its offensive rebounds percentage is low.

A team that plays good defense and has good capacity in ball handling, letting a low number of turnovers per possession, achieves the playoffs even if it keeps a low offensive rebounds rate.

30. If a team plays a defensive basketball and keeps the opponents to a low field goal percentage, it will achieve the playoffs even if its opponents often reach the foul line with a good free throw percentage.

A team that is good defensively and that forces the opponents to miss throws (or to shoot poorly), even if it could mean to make a lot of fouls, reaches the playoff and is able to strive for the title in post-season.

31. A newly promoted team, or a team with a medium roster stability, that allows the opponents to high field goal percentage and is at most on average on defensive rebounding, will finish the regular season in the middle of the ranking.

Newly promoted teams, or teams with medium roster stability, cannot reach more than a center-ranking position if they allow a lot defensively, letting the opponents to shoot from the field and to grab a lot of offensive rebounds.
33. If a team plays an offensive basketball but does not force a lot of turnovers, it will finish the regular season in the middle of the ranking, even keeping a high defensive rebounds rate. An offensive team, with a high percentage of defensive rebounds, ends the regular season in center class if the number of turnovers forced is, at most, on average.

33. If a team plays an offensive basketball, it is not able to get to the foul line (or keeps a low free throws rate) but, conversely, lets the opponents to do it, at most it can run for a middle ranking position. An offensive team that rarely gets fouls or has a low free throws percentage ends up in center class, especially if it lets the opponents to an average free throws percentage.

34. A newly promoted team, or a team with low roster stability, that shoots poorly and let the opponents to do it very well, ends in relegation class even if its defensive rebounds rate is high. A team which is newly promoted, or has a low roster stability, finds itself in an unfavourable position. This means that an effective defensive rebounding does not prevent it to finish in relegation, if the team keeps a low field goal percentage and its opponents are free to keep it high.

35. A team with low roster stability that lets the opponents a high field goal percentage, ends in relegation class even if it has average ball handling skills. Even if it allows an average number of turnovers per possession, a team ends the season in relegation class if its roster is not stable and so it is its defense, letting the opponents to make a lot of shots from the field.

36. A defensive team which is not able to prevent the opponents to scoring a lot, and besides it has a low field goal percentage, will end the season trying to avoid relegation. There are instances of teams that play better on defense then in offense, so they found themselves shooting poorly from the field, but paradoxically cannot defend the basket letting the opponents making a lot of field goals. These teams cannot aim for nothing better then avoiding relegation.

Summarizing, it is easy to say that teams should be consistent on their play. Indeed, teams that are strong in the factors that they decide to concentrate on are usually rewarded. For instance, it is relatively easy to reach the playoffs deciding to play a defensive basketball and concentrate on not letting the opponents to shoot easily from the field. Moreover, it seems that shooting percentage, both on offense and on defense, are the most important factors, and keeping them above average (or below, for \( f_{1\text{def}} \)) means a lot in terms of final rankig position.
These results seem to have some points in common and some differences with what has been expressed by the literature. Namely, what is in common is the relevance of certain aspects of the game, i.e. the shooting percentage, both in offense and in defense, above all. This reinforce also Berri’s idea of the importance of offensive and defensive rating, two stats that under this light assume a lot of meaning - anyway, not so much to become the only factors which is useful to look at. In fact, Rough Set Theory enhance the centrality of other factors that were poorly considered: an example is given by Implication 4, which reverses the scale of values of the most important factors set by standard literature. Here is the strength of this theory, the capacity to determine the importance of a set of condition attributes.
<table>
<thead>
<tr>
<th>Rule</th>
<th>IF (Condition attributes)</th>
<th>THEN (Decision attribute)</th>
<th>Coverage</th>
<th>Certainty</th>
</tr>
</thead>
</table>
| 25   | $f_{1}^{off} = \text{HIGH}$  
      | $f_{2}^{off} = \text{HIGH}$  | Title | 30.00% | 100.00% |
| 26   | $f_{1}^{off} = \text{HIGH}$  
      | $f_{1}^{def} = \text{MEDIUM}$  
      | $f_{4}^{def} = \text{HIGH}$  | Title | 25.00% | 100.00% |
| 27   | $f_{1}^{off} = \text{HIGH}$  
      | $f_{3}^{off} = \text{HIGH}$  
      | $f_{4}^{off} = \text{MEDIUM}$  | Title | 20.00% | 100.00% |
| 28   | $f_{4}^{off} = \text{HIGH}$  
      | $f_{1}^{def} = \text{HIGH}$  
      | Overall = DEF | Playoff | 20.00% | 100.00% |
| 29   | $f_{3}^{off} = \text{LOW}$  
      | $f_{2}^{def} = \text{HIGH}$  
      | Overall = DEF | Playoff | 20.00% | 100.00% |
| 30   | $f_{1}^{def} = \text{HIGH}$  
      | $f_{4}^{def} = \text{LOW}$  
      | Overall = DEF | Playoff | 25.00% | 83.33% |
| 31   | $f_{1}^{def} = \text{LOW}$  
      | $f_{3}^{def} \leq \text{MEDIUM}$  
      | Syn = MEDIUM or NEW | Center | 28.00% | 100.00% |
| 32   | $f_{2}^{def} = \text{MEDIUM}$  
      | $f_{3}^{def} = \text{HIGH}$  
      | Overall = OFF | Center | 24.00% | 100.00% |
| 33   | $f_{4}^{off} = \text{LOW}$  
      | $f_{4}^{def} = \text{MEDIUM}$  
      | Overall = OFF | Center | 16.00% | 100.00% |
| 34   | $f_{1}^{off} = \text{LOW}$  
      | $f_{1}^{def} = \text{LOW}$  
      | $f_{3}^{def} = \text{HIGH}$  
      | Syn = LOW or NEW | Relegation | 40.00% | 100.00% |
| 35   | $f_{2}^{off} = \text{MEDIUM}$  
      | $f_{1}^{def} = \text{LOW}$  
      | Syn = LOW | Relegation | 40.00% | 85.71% |
| 36   | $f_{1}^{off} = \text{LOW}$  
      | $f_{1}^{def} = \text{LOW}$  
      | Overall = DEF | Relegation | 26.67% | 100.00% |

Table 10: Main rules obtained with ROSE2
4.4 Strategies

What ROSE2 is able to calculate is also a set of indirect rules, i.e. implications that keep the decision attribute as starting point. Suppose you are a decision maker (let’s say the General Manager) of a basketball team competitor for Legabasket. Depending on how ambitious you are, and how strong your team is, you can develop a working plan that tells you where to concentrate: hereafter are presented some general strategies for your team in order to reach a specific ranking class. These strategies are later summarized in Table 11. As already specified, certainty of a direct rule becomes coverage of the indirect rule, while coverage of a direct rule becomes certainty of the indirect rule.

**Title**  If you are positive and ambitious enough to aim for the title, you’d better not only have a strong team, but also concentrate on some specific features. First of all, your team needs to have a high shooting percentage, let’s say at least 54% (two-pointers and three-pointers together). Independently from the total number of field goals attempted, your players should make them, and should care that they make on average more than the opponents.

Moreover, your team has to play a good defense, and specifically try to stop the opponents’ attempts to shoot from the field. This aggressive strategy could imply committing a significant number of fouls, but it does pay back in the long run. 40% of team that played this double strategy ended up in a good position to win the title.

**Playoff**  If you have playoff ambitions, but you don’t really care about the ranking position and you just want to end up among the first eight teams, you’d better train your team to play an effective defense. First thing to do is improve the capacity of blocking opponents’ shots, or at least play in such a way that they will be forced to make a number of shots lower than the league average.

Moreover, what emerges is that the majority of teams that played better on defense than they did in offense eventually reached the playoffs. That means that, respectively to league average, your team should concentrate on the defensive rating. This strategy has a coverage of 64.71% and a certainty of 55%.

**Center**  Lowering your ambitions, if you realize your team is not strong enough to compete for the playoff but neither to avoid to struggle for relegation zone, you may want to aim to a middle-ranking class. To do so, your team should focus on playing a purely offensive basketball. It’s the idea of counter attack: if you are an underdog, do your best to undermine opponents’ probabilities to win. And given that favorite teams would play defensively, attack them.
The opposite side of the same medal stands on the fact that focusing on offense would lead to a lack of some defensive aspects, i.e. showing a low capacity of stopping opponents’ shooters ($f1\text{def} = \text{LOW}$). This strengthen the hypothesis that teams that aim to a centre-ranking zone must concentrate on the offensive phase. This double strategy has a coverage of 50% and a certainty of 40%.

Relegation Clearly, relegation cannot be an aim for any team, despite its limits and weaknesses. 66.67% of teams with low capacity of stopping opponents' shooters and low roster stability ended up in relegation zone; it’s easy to say that if your team wants to avoid relegation, it should be at least on average on these factors. This means that a starting point in order to avoid relegation should be that of maintaining a certain degree of roster stability, in order to develop a good teamwork between your players. This rule has both coverage and certainty equal to 66.67%.

<table>
<thead>
<tr>
<th>Aim</th>
<th>Strategy</th>
<th>Coverage</th>
<th>Certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>$f1\text{off} = \text{HIGH}$</td>
<td>72.73%</td>
<td>40.00%</td>
</tr>
<tr>
<td></td>
<td>$f1\text{def} = \text{HIGH}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Playoff</td>
<td>$f1\text{def} = \text{HIGH}$</td>
<td>64.71%</td>
<td>55.00%</td>
</tr>
<tr>
<td></td>
<td>$\text{Overall} = \text{DEF}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centre</td>
<td>$f1\text{def} = \text{LOW}$</td>
<td>50.00%</td>
<td>40.00%</td>
</tr>
<tr>
<td></td>
<td>$\text{Overall} = \text{OFF}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relegation</td>
<td>$f1\text{def} = \text{LOW}$</td>
<td>66.67%</td>
<td>66.67%</td>
</tr>
<tr>
<td></td>
<td>$\text{Syn} = \text{LOW}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Main strategies obtained with ROSE2
5 Conclusions

The research implemented in this thesis has presented a new approach in the field of quantitative analysis of basketball team performance, in order to identify which are the main aspects a team should concentrate on to win its championship (or to achieve a specific ranking position). This method provides a new way of evaluating team performance, and it is mostly based on the use of Rough Set Theory - a mathematical and statistical tool to deal with datasets characterized by vague, incomplete or lacking information.

Thus, the first step was identifying, within the literature, the indexes that are more efficient in describing and summarizing how a team generally performs. These indexes were found in Oliver’s “eight factors” that have been proved to significantly describe how team perform over the season. Two more indexes were added (roster stability and overall playing tendency) in order to improve completeness. These ten indexes, conveniently classified, represented the conditional attributes for every team, along with the decisional attribute, i.e. final ranking classification. Then, software ROSE2 was used to find and study some certain decision rules, which were useful also in order to derive some possible management strategies.

One obvious expansion of this work is to enlarge the database, using data from more seasons. This could lead also to an interesting time series analysis, involving the opportunity to study how things changed over decades and to find common trends. Another possibility could be that of include leagues different from Legabasket. In this sense, it would be interesting to analyze the championship of the teams participating to Eurolega, so to understand the different paths teams have to follow in order to achieve it. Also, a different possibility could be that of implementing Dominance-based Rough Set Approach (DRSA): here, dominance relation substitutes indiscernibility relation on which the classical Rough Set Theory has been established. In effect, DRSA is an extension of RST for multi-criteria decision analysis.

Another possible expansion could be that of using different condition attributes, for example those derived from Berri’s work on evaluating team performance, or that of using a more specific classification method. In effect, it is difficult to find an equilibrium of the trade-off between consistency and certainty of the decision rules and the increasing specification of the classes of the decision attributes.
6 Acknowledgements

First of all, I would like to thank my supervisor, professor Paola Ferretti, for her kind willingness to help me and to support me in one of the toughest moments of my academic career. Her precious advice and brilliant ideas were extremely important, and the feeling of cooperation I have found with her represented a great incentive to develop the research the better I could. I will remember our serene and profound conversations for the rest of my life.

I would also like to thank the colleagues that shared with me all the good moments - and difficulties - we have lived in these past two years. It has been great to get to help each other, as well as having always someone to share projects and ideas with.

I would like to thank Andrea, Claudio, Tommaso and Daniele for their friendship and the awesome moments we have spent together since we met the first time. My life without you would have been way less interesting, and I wish you all the best.

Thanks to my australian friend Ben, his wife Eloisa and their beautiful families in Gold Coast and São Paulo, that were so kind to share with me such an important event as their wedding in Brazil. That is a breathtaking memory, together with an amazing journey, and I will never forget it. Also, I would like to thank Ben for his time and the dedication he has devoted to me. You have all my respect, and I can’t wait to pay you back.

Thanks to two of the souls that have had, and continue to have, the greatest impact on my life, Edoardo and Nicola. Your presence has taught me a lot, both in how to live my life and how to fight to get what I want. Thanks for always being at my side, for always being honest and open to each other. Our relationship has helped me to achieve goals that I thought I couldn’t achieve, and to trust myself more than I would usually do. This is such a great motivation, and I am always looking forward to the next time we see each other altogether.

Lastly, I would like to thank my family. For all the love I feel from my brothers, and all the support I get from my parents. Without you all this would not have been possible. This is why I entirely dedicate this work to you.
References


7 Appendix

Hereafter the list of teams related to each rule is presented, along with the season in which they made those results. Tables that contain rules are from Table 5 to Table 10, so the list of records is divided depending on the Table they belong to.

Table 5 (p. 44)  Rule 1: (Milano, 13/14) (Milano, 12/13) (Milano, 11/12) (Cantú, 11/12) (Siena, 09/10).
Rule 2: (Cantú, 13/14) (Sassari, 13/14) (Sassari, 11/12).
Rule 3: (Siena, 12/13) (Venezia, 12/13) (Treviso, 10/11) (Bologna, 09/10).
Rule 4: (Bologna, 11/12) (Treviso, 10/11) (Bologna, 09/10).
Rule 5: (Montegranaro, 13/14) (Bologna, 12/13) (Biella, 12/13) (Brindisi, 10/11) (Ferrara, 09/10).

Table 6 (p. 45)  Rule 6: (Cantú, 13/14) (Sassari, 13/14) (Siena, 11/12) (Sassari, 11/12) (Cantú, 10/11).
Rule 7: (Bologna, 11/12) (Pesaro, 11/12) (Roma, 09/10).
Rule 8: (Brindisi, 13/14) (Siena, 12/13) (Treviso, 10/11) (Bolona, 09/10).
Rule 9: (Varese, 13/14) (Avellino, 13/14) (Cremona, 11/12) (Biella, 11/12) (Roma, 11/12).
Rule 10: (Pesaro, 13/14) (Montegranaro 13/14) (Montegranaro, 11/12) (Teramo, 10/11).

Table 7 (p. 47)  Rule 11: (Sassari, 13/14) (Avellino, 10/11) (Cantú 09/10).
Rule 12: (Brindisi, 13/14) (Siena, 12/13) (Bologna, 09/10).
Rule 13: (Venezia, 13/14) (Caserta, 12/13) (Caserta, 10/11) (Teramo, 09/10).
Rule 14: (Bologna, 12/13) (Caserta, 11/12) (Monferrato, 11/12) (Brindisi, 10/11).
Rule 15: (Cremona, 13/14) (Biella, 12/13) (Ferrara, 09/10).

Table 8 (p. 48)  Rule 16: (Siena, 11/12) (Siena, 10/11) (Siena, 09/10).
Rule 17: (Pistoia, 13/14) (Venezia, 11/12) (Sassari, 10/11).
Rule 18: (Roma, 13/14) (Venezia, 11/12) (Sassari, 10/11).
Rule 19: (Brindisi, 12/13) (Cremona, 11/12) (Roma, 10/11).

Table 9 (p. 49)  Rule 20: (Siena, 11/12) (Sassari, 11/12) (Siena, 10/11) (Siena, 09/10).
Rule 21: (Avellino, 12/13) (Cremona, 12/13) (Roma, 11/12) (Caserta, 10/11) (Pesaro, 09/10) (Teramo, 09/10).
Rule 22: (Brindisi, 13/14) (Siena, 12/13) (Treviso, 10/11) (Bologna, 09/10).  
Rule 23: (Pesaro, 13/14) (Montegranaro, 13/14) (Montegranaro, 11/12) (Teramo, 10/11).  
Rule 24: (Bologna, 12/13) (Biella, 12/13) (Monferrato, 11/12) (Ferrara, 09/10).  

Table 10 (p. 54)  
Rule 25: (Varese, 12/13) (Sassari, 12/13) (Siena, 11/12) (Sassari, 11/12) (Siena, 10/11) (Siena, 09/10).  
Rule 26: (Cantú 13/14) (Sassari, 13/14) (Sassari, 12/13) (Sassari, 11/12) (Cantú, 09/10).  
Rule 27: (Milano, 13/14) (Milano, 12/13) (Milano, 11/12) (Cantú, 11/12).  
Rule 28: (Pistoia, 13/14) (Bologna, 11/12) (Pesaro, 11/12) (Bologna, 10/11).  
Rule 29: (Brindisi, 13/14) (Siena, 12/13) (Bologna, 09/10) (Roma, 09/10).  
Rule 30: (Milano, 13/14) (Brindisi, 13/14) (Reggio Emilia, 13/14) (Siena, 12/13) (Reggio Emilia, 12/13) (Venezia, 12/13).  
Rule 31: (Varese, 13/14) (Avellino, 13/14) (Brindisi, 12/13) (Cremona, 11/12) (Biella, 11/12) (Varese, 09/10) (Cremona, 09/10).  
Rule 32: (Avellino, 12/13) (Cremona, 12/13) (Roma, 11/12) (Caserta, 10/11) (Pesaro, 09/10) (Teramo, 09/10).  
Rule 33: (Varese, 13/14) (Brindisi, 12/13) (Cremona, 11/12) (Roma, 10/11).  
Rule 34: (Pesaro, 13/14) (Bologna, 12/13) (Caserta, 11/12) (Monferrato, 11/12) (Teramo, 10/11) (Brindisi, 10/11).  
Rule 35: (Treviso, 09/10) (Montegranaro, 13/14) (Bologna, 12/13) (Biella, 12/13) (Biella, 10/11) (Teramo, 10/11) (Ferrara, 09/10).  
Rule 36: (Bologna 12/13) (Biella 12/13) (Monferrato, 11/12) (Ferrara, 09/10).