Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE

Corso di Laurea in Informatica

Minutes played by Under 21 players in Serie A: a descriptive analytics study

Relatore:
Chiar.mo Prof.
Marco Roccetti

Presentata da:
Francesco Piconi

IV Sessione
Anno Accademico 2022/2023

Sommario

Storicamente, la Serie A è sempre stata considerata un campionato difficile per la crescita dei giovani talenti, in quanto offre poche opportunità ai giocatori emergenti di mettere in mostra le proprie capacità e potenzialità. Questo studio esplora proprio questo aspetto del massimo campionato di calcio italiano: i minuti giocati dai giocatori under 21 e quelli giocati dai giocatori under 21 italiani. L'obiettivo è analizzare le tendenze che si sono sviluppate nel corso delle varie stagioni, facendo luce sull'evoluzione e l'impatto dei giovani talenti in uno dei principali campionati di calcio europei. Lo studio metterà in luce in che modo i cambiamenti normativi hanno influenzato positivamente o negativamente le opportunità concesse alle giovani promesse. Questa ricerca esamina meticolosamente i dati nel corso delle stagioni, analizzando i minuti giocati dai giocatori under 21 e e del sottoinsieme dei calciatori italiani under 21. Questi dati vengono utilizzati per determinare l'influenza su due diversi fenomeni. Nello specifico, i minuti giocati dagli under 21 sono correlati e causano i ricavi delle squadre di Serie A, mentre i minuti giocati dagli under 21 italiani non causano i risultati della Nazionale italiana. I risultati di questa ricerca hanno lo scopo di fornire informazioni preziose per tifosi di calcio, analisti e stakeholder, offrendo una comprensione approfondita del fenomeno di crescita e sviluppo dei giovani talenti nel massimo campionato di calcio italiano.

Abstract

Historically, Serie A has always been regarded as a difficult league for the development of young talent, as it offers few opportunities for emerging players to showcase their skills and potential. This study explores a critical aspect of Italy's top football league: the minutes played by under 21 players and, more specifically, by Italian under 21 players in Serie A. The goal is to analyze the trends and patterns that have developed over the course of the various seasons, shedding light on the evolution of the role and impact of young talent in one of Europe's leading soccer leagues. The study will relate how regulation change, has positively or negatively affected the opportunities allowed to promising young players. This research meticulously examines data over the years, analyzing the minutes played by under 21 players and the subset of Italian under 21 players. These data are used to determine the influence on two different phenomena. The minutes played by the under 21 are related and cause the revenue of the Serie A teams, while the minutes played by the Italian under 21 do not cause results of the Italian National team. The results of this research aim to provide valuable information for football fans, analysts and stakeholders, offering a nuanced understanding of the approach of the Italian Top League of football to nurture and integrate young talents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Serie A has always been considered a challenging league for young talent development, with limited opportunities for emerging players to showcase their skills and potential. This study delves into a critical aspect of this developmental landscape: the minutes played by under 21 players and, more specifically, Italian under 21 players in Serie A. The objective is to dissect the trends and patterns that have unfolded over various seasons, shedding light on the evolving role and impact of young talents in the top Italian football league.
The Serie A has long been renowned for its tactical sophistication and competitive nature, making it an ideal testing ground for burgeoning talents. As clubs navigate the delicate balance between immediate success and long-term sustainability [1], the role of young players becomes increasingly pivotal. Understanding the patterns of utilization and progression of under 21 talents provides a unique lens through which we can observe the league's commitment to youth development and the broader implications for Italian football on the international stage. The research aims to shed light on the possible relations of the minutes played by the under 21 on two very different aspects: the revenues in the transfer market of Serie A teams and the results of the Italian national team.

1.1 Literature review

Many papers in the literature analyze the minutes granted to under 21 players. Often the goal is to look for a correlation between the introduction of a new regulation and attendance in matches [2] or minutes given to young players. Since the 1990s, European football has been shaped by three significant regulations affecting clubs and league competitions, with two key directives imposed by UEFA, the governing body of the football in Europe. The first was the 1995 Bosman ruling, a landmark decision by the European Court of Justice [3] on freedom of movement for workers. Subsequently, UEFA introduced two impactful regulations: the Home Grown Player Rule in 2006 and Financial Fair Play (FFP) in 2011. These regulations directly impact player development strategies and migration in European football. The December 1995 ruling of the European Court of Justice[3](Bosman rule) resulted in the liberalization of professional sportsmen's migration within the EU and the abolition of transfer fees after contracts expired. Thus, there has been a transfer of the ownership rights of the tag from the clubs to the players themselves. In addition, the elimination of foreign player quotas for EU clubs could lead to a significant migration of players and thus a reallocation of players. The work presented by Marcén, M.[4] analyzes whether the Bosman ruling had a negative impact on the number of national players participating in their respective national leagues by analyzing the case of the Spanish football league LaLiga; the work show the negative impact on the number of minutes played by Spanish players. Furthermore, around this new rule there was much scepticism due to the fear that clubs would no longer invest in their youth sectors [5] to grow new players consequently penalizing the level of the national teams. Even UEFA's chief executive Lars-Christer Olsson said [6] in 2005, ten years after the Bosman ruling, that the decision ruined football values and led to the abandonment of national talent. In fact as analyzed by Binder et. al. [7], the liberalization of transfers reduces the chances given to young players to establish themselves in the top leagues. The possible implications of lowering the level of national teams is refuted

by the same author in his paper with very small effects.
The Home Grown Player Rule aimed to promote the development of local talent, by ensuring that a minimum number of players in each squad were home-grown. The purpose was to have from the 2008-2009 season in the 25 players registered for each team, 4 grown in their home league and 4 grown in their own team. This rule was mandatory for teams taking part in European competitions, but since 2011-2012 also teams participating in the top Italian football league Serie A [8] must comply with the Uefa home-grown rule. The work presented by Bullogh et al.[9], analyzes which effects the home-grown rule has had on the minutes allowed to native players in major European leagues, with a particular focus on under 21 players. The sample of seasons analyzed stops at 2015-2016, showing slight variations in the Italian Serie A league, and a significant increase in minutes played by under 21 players in the German Bundesliga.
A very similar home-grown rule adopted by UEFA is the under 21 rule introduced by the Royal Belgian Football Association; the rule requires Belgian domestic division teams to include at least two players under the age of 21 in match selection(starters and substitutions during the match). The analysis by Roel Vaeyens et. al.[2] concludes that the rule proposed by the Belgian federation has failed to increase playing opportunities for young players. Thus, not all rules created to protect the opportunities granted to under 21 football players yield the desired effects. On the contrary, there is an example of the so-called Financial Fair Play introduced by UEFA focused on ensuring financial stability of major European clubs, and that has generated more investment in the development of young players. In fact as argued by Regoliosi C.[10], "the profitability of Italian football clubs depends on their ability to manage a Business model "player development" : that is, discovering, signing, developing and selling talented young players, despite the fact that they cannot achieve significant sporting results at the moment. In this way Italians clubs will be able to maintain their sporting and financial results while observing the new regulation imposed by UEFA on balanced

budgets. Starting from the idea presented by Religiosi C.[10] in his work, the possible correlation between the minutes granted to the under 21 players and the revenues obtained by the Italian teams from the transfer market was analyzed.

1.2 Reasons for the research

The study presented has as its objective and purpose to debunk one of the clichés about the Top Italian football championship: the Italian teams do not play young players. Understanding the distribution of playing time for under 21 players can shed light on how Serie A clubs prioritize the development of young talent. Analyzing how changes in Serie A rules affect the playing time of under 21 players, can shed light on the impact of these rules on the development of young talent. This could be of interest to football clubs, academies and governing bodies looking to improve youth development programmes. The minutes played by under 21 players can reveal information about how teams adapt to rule changes and whether they are more or less likely to give opportunities to young players. Additionally, comparing the minutes played by under 21 players before and after the rule changes can reveal trends and patterns. This comparative analysis can help identify whether rule changes have had a significant impact on the inclusion of young players in Serie A matches.

Thus, the thesis will focus on several research questions:

RQ1: examine the influence that rule changes in Serie A have had on minutes played by under 21 and Italian under 21 players, with a focus in the last period looking for a positive trend

RQ2: analyze the correlation and a causality between the profits made in football market sessions by Italian teams and the minutes granted to under 21 players [10]

RQ3: investigate the correlation and causality between the minutes played by Italian under 21 players and the results achieved by both the Italian national team and the Italian under 21 national team

Chapter 2

Data and Methods

The methodology and data examined in this study involve a comprehensive analysis of minutes played by under 21 players, with a specific focus on Italian under 21 players, in Serie A over several seasons, with different effects produced. The study is divided into two main parts:

- Quantitative and Qualitative Analysis: this is the statistical analysis of minutes played by under 21 players and Italian under 21 players in Serie A. The presence of trends and patterns over seasons, with the impact of regulatory changes on opportunities for young players

- Correlation and Casualty Analysis: the data, are then used to determine the influence on two different phenomena. The relationship between minutes played by under 21 players and revenues from the sale of under-23 players by Serie A teams. The impact of minutes played by Italian under 21 players on the results of the Italian national team and the Italian under 21 national team.

2.1 Data

The data for this thesis were obtained from different sources. The information about the under 21 players were scratched from *transfermarkt*, a German-based website that has footballing information, such as scores, results, statistics, transfer news. The Serie A seasons analyzed cover the period from 1967-1968 to 2022-2023. Only players under 21 years old before the start of the football season were examined. The information for under 21 players are: IdPlayer, Nationality, MinutesPlayed, Age, Team. From this data, the minutes played by under 21 players are counted for each year, then split by Italian players, oriundi (Italians with dual nationality) players and foreign players. The minutes played by under 21 players were then normalized according to the total number of minutes available in each season (over the years there have been three changes in the number of teams participating in the championship). The result of normalization is then the percentage of minutes given to under 21 players per Serie A season. The decision to normalize the data was motivated by the varying number of teams in different seasons: 1967-1988: 16 teams, 1988-2004: 18 teams, 2004-present: 20 teams. This normalization aims to ensure consistency and uniformity in the minutes played by the under21s, allowing accurate comparison across seasons.

Profits earned by Serie A teams have been collected from *transfermarkt* since the 1993 season, as data prior to 1990 are incomplete and do not allow for an objective analysis of the trend. Profits were filtered by taking into account only those produced by the sale of under-23 players. This choice is motivated by the possibility of finding a correlation between the sales of under-23 players and minutes played by under 21 players, with an analysis shifted over time. In fact, it is plausible to assume a sale of young players after a few seasons. Therefore, the influence of minutes played by under 21 players shows the effects on revenues from disposals in later years, when the players are over 21 years old.

Two different approaches were chosen to evaluate the performance obtained by the Italian national football team and the Italian national under 21 team. FIFA regularly assigns scores to national football teams, allowing us to objectively compare the relative strengths of national teams active in the international arena. This score is called the FIFA Ranking[11] and is based on

the results obtained during the year, based on the number of wins, losses and draws obtained by national teams. This score is also influenced by the results obtained during international events such as the World Cup. The FIFA Ranking has been calculated since 1992 and from the official FIFA website you can access the score of all the teams. Being a score that is increased over the years it is difficult to make a comparison with previous years. To solve this problem, scores of the top 50 national football teams were obtained for each year. In order to achieve a comparable score over time, the Italian national team's score was divided by the average score of the first 50 teams. To assess the performance of the Italian under 21 national team, the total matches played each year from 1993 were obtained. For each year was then calculated the average of points obtained (win 3 points, draw 1 point, defeat 0 points). It was also given weight to the results obtained by the national under 21 at the European Championship that is held every two years. Then based on the placement obtained during the tournament the following points are awarded: qualification 0.25, quarter-final 0.5, semi-final 0.75, final 1, victory 1.25.
The data were obtained by scraping techniques using Python and in particular the Beatiful Soup library that simplifies scraping information from web pages. The pandas and numpy libraries were then used to clean and view the data, and finally create the dataset to perform the analysis. The code is available on the project's Github page. [12]
The collected data were merged using the year as an index obtaining a dataset consisting of 7 columns. A preview of the result is shown in Figure 2.1.

	Minutes_played_under21	Minutes_played_italians	Minutes_played_foreigners	Fifa_points	Points_naz_u21	revenue
2018	8.221823	2.669724	5.552100	1.064686	1.350	5.489717
2019	6.539474	2.618820	3.920654	1.133279	2.555	6.272201
2020	6.049575	2.003721	4.045853	1.159438	3.000	5.480558
2021	5.818846	2.303961	3.514886	1.386975	2.475	5.337017
2022	6.184476	3.170920	3.013557	1.312050	1.750	3.891005

Figure 2.1: Final dataset

The football seasons straddle two calendar years, so they are indentified as $startyear - endyear$, e.g. 2022-2023. Minutes played represent the percentage of minutes allowed during the Serie A season, normalized as described above. For simplicity then, only the starting year of the football season is considered, so in the above example the season is considered as 2022.
The minutes played were divided into those played by players with Italian nationality and those played by foreign players.
Revenues of Serie A teams by the sale of under-23 players were preprocessed following two steps. They were first adjusted to obtain the real value, that is, the nominal value adjusted with inflation. This is done using the Consumer Price Index obtained from The World Bank[13], which measures the overall change in consumer prices based on a representative basket of goods and services over time. Then, given the exponential nature of the value of transfers in the soccer market, logarithm was applied, a technique that is often used to normalize economic data.

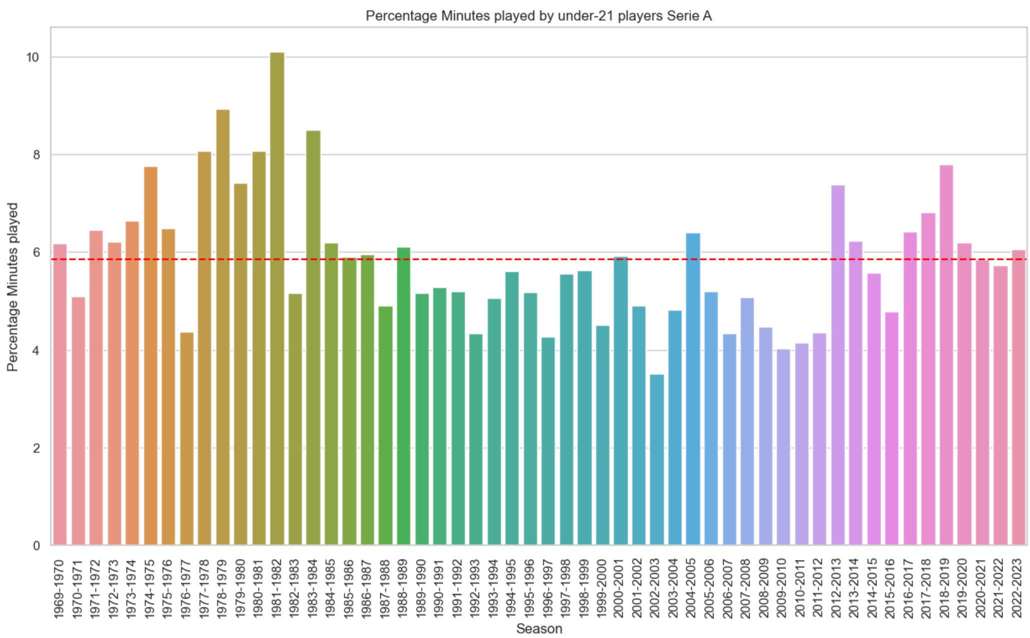

Figure 2.2: Percentage Minutes played by under 21 players

Figure 2.2 shows the percentage of minutes played by under 21 players over the various seasons, with the average standing at 6.20% and standard deviation of 1.33.

To show the distribution of minutes played by under-21 players, and to obtain a more detailed visualization, data were divided into three groups: players of Italian nationality, Italian players with dual nationality(Oriundi), and finally foreign players. This distribution of data can be seen in Figure 2.3. This classification was adopted for demonstration purposes only, and in the subsequent work the classification shown in Figure 2.1 was preferred, thus with two groups: players of Italian nationality and foreign players.

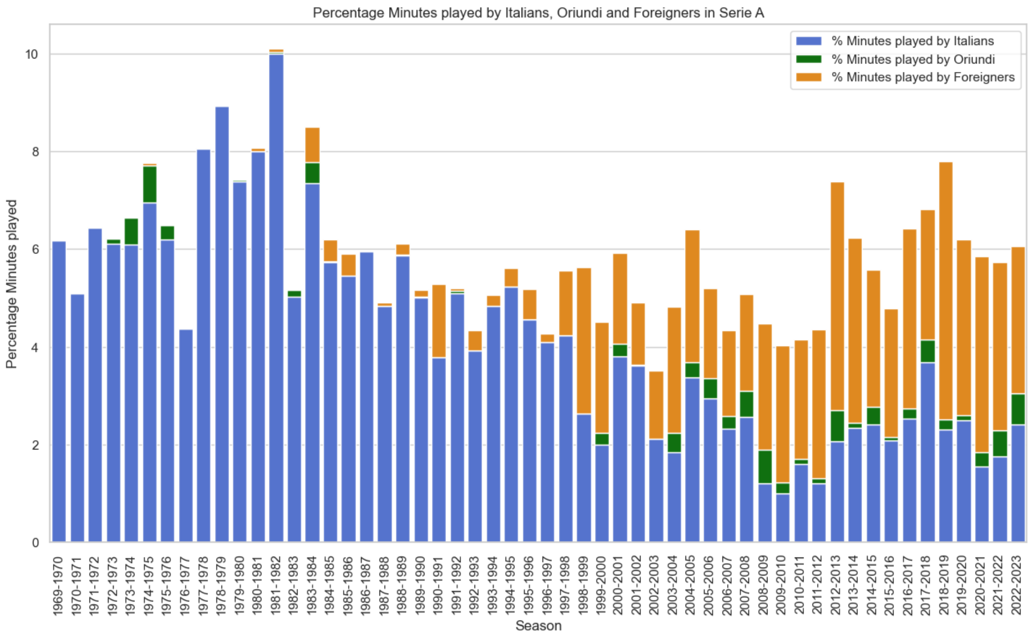

Figure 2.3: Minutes played by Italian, foreign and oriundi under 21 players

To obtain a narrower view of the phenomenon, a grouping was performed every eight years, creating two graphs on the figure 2.4. The graph on the left shows the trend in the percentage of minutes played by under 21 players grouped every 8 years. The graph on the right instead shows the same data but divided between foreign footballers and Italian footballers.

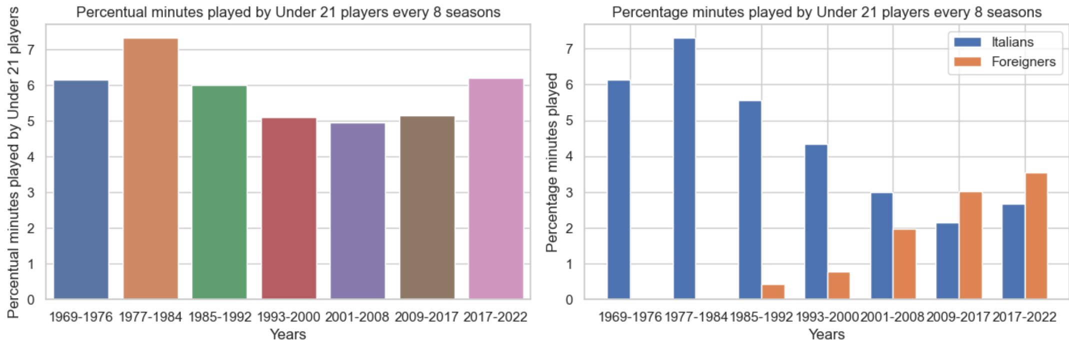

Figure 2.4: Minutes played by under 21 players every 8 years

2.2 Methodologies

To understand how the percentages of minutes given to under 21 players have varied over the years, and whether there has been a positive increase in recent years, a time series analysis was performed.
The analysis integrated with change point detection allows us to determine when changes in the behavior of a time series are present and whether these coincide with the introduction of new rules in the Serie A regulations.
A ranks-based test (Mann-Kendall test) will be used to statistically determine the presence of positive or negative trends found as a result of regulation changes. To complete the study concerning the presence of trends, unit root-based tests (Augmented Dickey Fuller test and Kwiatkowski-Phillips-Schmidt-Shin test) will be introduced to determine the stationarity of the series. In order to establish the correlation and casuality between the minutes played by under21 players and the revenues obtained in the soccer market by Italian teams, two different techniques were applied. Initially, the relationship between the two variables was analyzed, establishing the Pearson correlation and in particular the cross-correlation that is useful as it identifies whether values in one series are predictive of future values in another series. Then, the existence of a cause-and-effect relationship between the variables was tested with Granger causality.

2.2.1 Time series

Time series can be defined as a series of data collected and analyzed in temporal order, in equally spaced time periods (e.g., by year, month, day, hour). The only independent variable in time series methods is time. Time series methods are based on the laws and techniques of statistical science, which can be defined as the branch of mathematics that deals with the collection, classification, analysis and interpretation of numerical data describing a phenomenon. Time series methods do not aim to explain or search for the causal structure of the problem but rather detect its trend and evolution over time.[14]They only give the possibility to understand, recognize and describe in the form of an analytical equation the evolution of a phenomenon over time. A time series can be written as:

$$y_t = S_t + T_t + R_t \tag{2.1}$$

and composed with following components :

- trend T_t, is non-stationary and reflects and incorporates the general characteristics and trend (increasing, decreasing, etc.) of the time series.

- seasonal S_t, component is a pattern of data that repeats after a period of days, weeks, months or quarters. It is also a non-stationary component and represents fluctuations around the curve of the trend component.

- residual R_t, also known as error term, disturbance or noise is an irregular component and has no discernible pattern, and reflects the effects on demand of known or unknown variables or factors that we ignore or are unable to measure and consider.

The graph in Figure 2.5 shows the time series of the percentage of minutes played by under 21 players in the Serie A league from the 1967-1968 season to 2022-2023 with trend component.

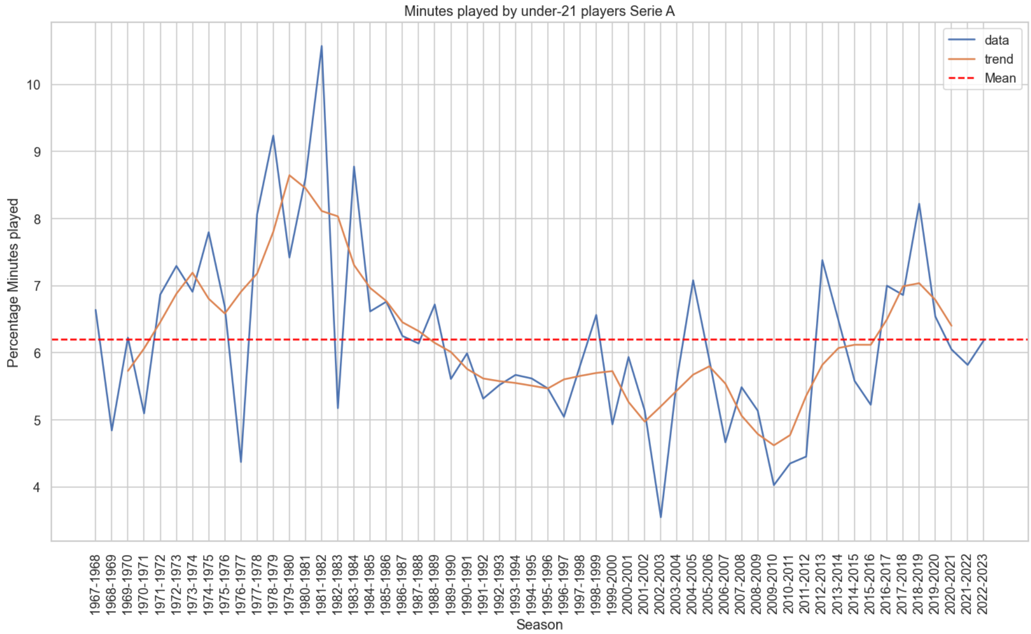

Figure 2.5: Time series of Minutes played by under 21 players

Time series decomposition involves breaking a series down into its components. Decomposition is primarily used for time series analysis and as an analysis tool it can be used to inform forecasting models about the problem. It provides a structured way of thinking about a time series forecasting problem, both in general terms of modeling complexity and in specific terms of how to best capture each of these components in a given model.

2.2.2 A Changepoint Detection Method

From the analysis of the components of a time series it is possible to determine change points, i.e. substantial changes in the trend of the data. The idea behind all these methods is, in general, to calculate some sort of measure of discrepancy between different parts of the observed time series. If the two parts do not have a significant change (for example in the mean or variance, or both) this discrepancy factor will be low, otherwise a peak will occur and a change point will be precisely identified. To find the change points in the time series represented graph of Figure 2.5, we used the change point selection method BEAST(Bayesian Estimator of Abrupt change, Seasonality, and Trend) described in [15]. BEAST is a Bayesian regression method designed to isolate periodic and trend signals from a time series and to pinpoint abrupt shifts in the two isolated signals. BEAST extends conventional trend analyses by fitting both linear and nonlinear trends and disentangling trends from seasonality. It employs Bayesian model averaging, embracing all candidate models rather than selecting just one, and has the capability of inferring nonlinear dynamics, distinguishing it from existing trend analysis methods. The model considers a time series $D = \{t_i, y_i\}_{i=1,....,n}$, to be composed of three components, seasonality, trend, abrupt changes, plus noise:

$$y_i = S(t_i; \Theta_S) + T(t_i; \Theta_T) + \varepsilon_i \tag{2.2}$$

Here, we assume the noise ε to be Gaussian with a magnitude of σ, capturing the remainder in the data not explained by the seasonal $S(\cdot)$ and trend $T(\cdot)$ signals. Following the common practice, general linear models are adopted to parameterize $S(\cdot)$ and $T(\cdot)$. Abrupt changes (i.e., changepoints) are implicitly encoded in the parameters Θ_S and Θ_T of the seasonal and trend signals. By analogy to linear regression, the time t and data y are independent and dependent variables, respectively; Θ_S and Θ_T are parameters to be estimated from the data D . Specifically, the seasonal signal S(t) is approximated as a piecewise harmonic model, defined with respect to p knots in time at $\xi_k = 1, \dots p$. These knots divide the time series into $(p+1)$ intervals $[\xi_k, \xi_{k+1}]$, $k = 0, \dots, p$

$$S(t) = \sum_{l=1}^{L_k} \left[a_{k,l} \cdot \sin\left(\frac{2\pi l t}{P}\right) + b_{k,l} \cdot \cos\left(\frac{2\pi l t}{P}\right) \right] ; \xi_k \leq t \leq \xi_{k+1}, k = 0, ..., p \tag{2.3}$$

Here, P is the period of the seasonal signal (i.e., one year in our cases); L_k is the harmonic order for the k-th segment and is an unknown segment specific parameter; and the coefficients $\{a_{k,l}, b_{k,l}\}_{l=1,\dots,L_k}$ are the parameters for sins and cosines. The trend $T(t)$ is modeled as a piecewise linear function with respect to m knots at $\tau_j = 1, \ldots m$, which divide the time span into (m + 1) intervals $[\tau_j, \tau_j + 1]$, $j = 0, \ldots, m,$.

$$T(t) = a_j + b_j t \qquad (\tau \leq t < \tau_{j+1}, j = 0, ..., m) \tag{2.4}$$

where a_j and b_j are coefficients of linear functions. The number of change points and their corresponding time are unknown parameters of τ_j. The related parameters can be reclassified into two groups: model structure parameters M and segment-specific coefficient parameters β_M:

$$\begin{aligned} M &= \{m\} \cup \{\tau_j\}_{j=1,\ldots,m} \cup \{p\} \cup \{\xi_k\}_{k=1,\ldots,p} \cup \{L_k\}_k = 0, ..., p \\ \beta_M &= \{a, b_j\}_{j+0,\ldots,m} \cup \{a_{k,l}, b_{k,l}\}_{k=0,\ldots,p;\ l=1,\ldots,L_k} \end{aligned} \tag{2.5}$$

The model structure M includes the number and corresponding time of seasonal and trend change points, and the seasonal harmonic order; β_M represents the coefficients of the seasonal and trend displacement sectioned model.

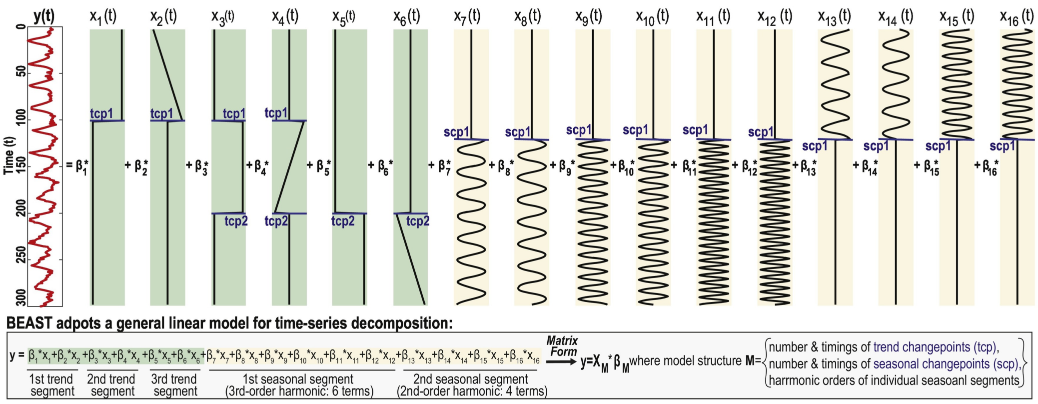

Figure 2.6: BEAST time series decomposition [15]

Considering now equation 2 we will have:

$$y(t_i) = x_m(t_i)\,\beta_m + \varepsilon_i \tag{2.6}$$

Identifying an optimal model structure M for our problem is analogous to choosing the best subset of variables for simple linear regression. Once a model structure M is selected, its coefficients B are straightforward to estimate.The problem is circumvented by resorting to Bayesian inference, to obtain the best values of these parameters, and their posterior probability distributions $p\left(\beta_M, \sigma^2, M| \left\{Y_t\right\}_t = 1, ..., n\right)$. But it is analytically intractable, so MCMC sampling is used to generate a realization of random samples for posterior inference. The MCMC sampling algorithm used is a hybrid sampler that embeds a reverse jump MCMC sample are simulated by reversible-jump Markov Chain Monte Carlo (MCMC) sampling. The MCMC algorithm generates a chain of posterior samples of length N $\left\{\beta_M{}^{(i)}, \sigma^{2(i)}, v^{(i)}, M^{(i)}\right\}_{i=1,...,n}$ (v is an extra dispersion hyperparameter introduce into it to further reflect the vague knowledge of the magnitude of model coefficients β_M). The chain captures all the information essential for inference of land dynamics, including trends, seasonal variations, and abrupt changes. Combining the individual estimates provides the final Bayes model averaging (BMA): $\hat{y}(t) \approx \sum_{i=1}^{N} x_M{}^{(i)}(t)\beta_M{}^{(i)}/N$. From the sampled model structure M the mean total numbers of trend and seasonal changepoints can be estimated, and the exact timings or an interval at which the trend or seasonal changepoints occurred for each of the N sampled models. Likewise, given an estimated changepoint, it's possible derive its credible interval.

The information estimated by BEAST not only includes seasonal and trend signals, seasonal and trend change points (scp or tcp), and the harmonic orders of individual seasonal segments. BEAST also provided a number of useful diagnostic statistics, such as, the confidence intervals of the estimated signals, the probability of observing a *scp* or *tcp* at a given time, the probability distribution of the total number of *scp* or *tcp*, and the probability of having a positive/negative change in the trend.

The BEAST algorithm was applied assuming the absence of seasonality in the data, so only the presence of trends will be analyzed.

2.2.3 Evolution of Serie A rules

Over the years there have been substantial changes in the regulations of the top Italian football league that may have affected the minutes allowed to under 21 players. They can be divided into four moments:

- In 1967, the leadership of Serie A made a pivotal decision attributing the continuous defeats of the national team to the lack of quality generational turnover. The blame was placed on foreign players in the league. Consequently, starting from the 1967-1968 season, the registration of foreign players was halted, leading to an inevitable surge in the involvement of Italians under 21 players across teams. The substantial minutes afforded to promising young talents between 1969 and 1980 can be attributed to this regulation imposed by the Serie A League. From the 1980-1981 season onwards, there was a gradual reintroduction of foreign players eligible for registration by teams. The number increased to 2 in the 82/83 season and further to 3 in the 87/88 season.

- significant shift that occurred from the 96/97 season onwards, the first season following the total revolution in European football brought about by the Bosman ruling. This ruling applied the principles of free movement of workers within the European Union to football, abolishing the limits on foreign players imposed on Serie A teams. In the latter part of 1995, the European Court of Justice (ECJ) [3] issued a groundbreaking verdict in the Bosman case, grounded in what is now Article 45 of the Treaty on the Functioning of the European Union (TFEU). This article outlines the rights of European Union (EU) nationals to engage in work on a non-discriminatory basis across any member state. The pivotal Bosman ruling had two key components. Firstly, it established that out-of-contract players who are EU nationals possess the right to negotiate their contracts with new clubs within the EU, devoid of their current clubs imposing a transfer fee as a prerequisite for their move. This perspective deemed the transfer system as an unwarranted constraint on the exercise of free movement rights. Secondly, the ruling declared the maximum limit of fielding three foreign players in a team as an unlawful restriction on freedom of movement and a violation of European discrimination law, given that players from EU member states were subjected to the same treatment as non-EU foreigners.

- the introduction of the Homegrown Player Rule [16] by UEFA in the 2008-2009 season. This regulation was implemented with the aim of fostering the development of local talent and nurturing a stronger connection between football clubs and their communities. This rule required having a minimum of eight home-grown players on the 25-man team. UEFA defines locally-trained or 'homegrown' players as those who, regardless of their nationality, have been trained by their club or by another club in the same national association for at least three years between the age of 15 and 21. Up to half of the locally-trained players must be from the club itself, with the others being either from the club itself or from other clubs in the same association. Since the 2011-2012 season, all teams participating in Italy's top Serie A football league [8] must comply with the Uefa rule on home-grown players.

- Financial Fair Play (FFP), originally introduced in 2009 by UEFA is based on the concept of financial sustainability to decrease club losses. In order to decrease operating expenses, many clubs began to invest in their youth sectors, from which they drew to complete their rosters and in many cases earn big money through the sale of promising home-grown youngsters.FFP mandates that clubs must break even, and at the same time it exempts a number of costs from the calculation like the full cost of youth development. This choice was voluntarily followed by UEFA to allow clubs to invest in youth sectors without having to worry about the budget.

2.2.4 Mann-Kendall test for monotonic trends

To statistically validate the presence of trends around the changepoints, the Mann-Kendall test for trends was used. The Mann-Kendall test for trends (sometimes called the MK test) is used to analyze time series data for consistently increasing or decreasing trends (monotonic trends). It is a nonparametric test, which means that it works for all distributions, and is particularly suitable for situations where the data may not follow a normal distribution as in our case.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}(x_j - x_i) \quad \mathrm{sgn}(z) = \begin{cases} 1 & \text{se } z > 0 \\ 0 & \text{se } z = 0 \\ -1 & \text{se } z < 0 \end{cases} \tag{2.7}$$

n is the length of the time series, $x_i$and x_j are the observations at times i and j, respectively, where $i < j$. With variance given by:

$$var(S) = \frac{n(n-1)(2n+5)}{18} \tag{2.8}$$

The significance of trends is tested with the standardized test statistic Z:

$$Z = \frac{S}{\sqrt{Var(S)}} \tag{2.9}$$

If the data has a serial correlation, it could affect in significant level (p-value). It could lead to misinterpretation. To overcome this problem, researchers proposed several modified Mann-Kendall tests, and the one proposed by Hamed and Rao [17] was selected to perform the work(Hamed and Rao Modified MK Test).

Therefore, the hypothesis test and the two hypotheses are as follows:

- **Null Hypothesis H0**: There is no trend present in the data
- **Alternative Hypothesis H1**: A trend is present in the data

2.2.5 Tests of Stationarity

Stationarity in a time series is characterized by constant statistical properties and the absence of seasonality. A time series is deemed stationary when both the mean and standard deviation remain constant, and there is no discernible trend or seasonality in the data. On the contrary, a non-stationary time series exhibits changing statistical properties over time, typically accompanied by the presence of trend and seasonality components. In such cases, the mean, standard deviation, and patterns within the time series may vary, making it essential to account for these dynamic characteristics in analyses. There are some test belongs to a category called *Unit Root Test*, which is the proper method for testing the stationarity of a time series. A unit root implies that a time series exhibits a stochastic or random trend, and its statistical properties do not remain constant over time. Technically, a unit root is said to exist in a time series if the value of $\alpha = 1$ in the equation:

$$Y_t = \alpha Y_{t-1} + X_e \tag{2.10}$$

where, Y_t is the value of the time series at time t and X_e is an exogenous variable (a separate explanatory variable, which is also a time series). The presence of a unit root means the time series is non-stationary Two common tests used to assess stationarity based on Unit Root Test are the Augmented Dickey-Fuller Test(DF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test.

Augmented Dickey Fuller test(ADF)

The Dickey-Fuller (DF) test was developed by Dickey and Fuller[18]. The Dickey-Fuller test is a unit root test that tests the null hypothesis that $\alpha = 1$ in the following model equation:

$$y_t = \alpha + \beta t + \gamma y_{t-1} + \phi \Delta y_{t-1} + \epsilon_t \tag{2.11}$$

where, ΔY_{t-1} is the first difference of the series at time $t-1$, γY_{t-1} is the lag 1 of time series, and ϵ_t is the error term. The intuition behind the test is that if the series is characterised by a unit root process then the lagged level of the series will provide no relevant information in predicting the change in besides the one obtained in the lagged changes. Therefore, Dicky and Fuller augment the equation with higher-order lags to capture the higher-order autocorrelation, known as the Augmented Dickey Fuller (ADF) test in the time series literature. The null hypothesis of the ADF test is that there is a unit root in an AR model, which implies that the data is nonstationary. The alternative assumption is generally stationarity, but it may be different depending on the version of the test used.

Kwiatkowski-Phillips-Schmidt-Shin Test(KPSS)

The KPSS test was created to integrate unit root tests because they have a lower power compared to trend processes that are close to the unit root and long-term. Unlike the Dickey-Fuller Test, the KPSS Test is designed to identify trends in time series data. It tests whether a time series is stationary around a deterministic trend. The null hypothesis of the KPSS Test is that the time series is stationary around a deterministic trend. Whereas in ADF test, it would mean the tested series is stationary.

Both tests are useful in determining the stationarity of a time series, and it's a good idea to use them together to get a more complete picture of the properties of the time series.

2.2.6 Correlation analysis

In statistics, correlation or dependence is any statistical relationship, causal or not. Correlations are advantageous because they can reveal a predictive relationship that can be utilized in practice. In statistics, correlation means a relationship between two variables such that each value of the former corresponds to a value of the latter, following a certain regularity. Correlation does not depend on a cause-and-effect relationship as much as on the tendency of one variable to change as a function of the other. Correlation is described by a correlation coefficient, ranging from -1 to +1 and denoted by ρ. Statistical significance is indicated by a p-value, which is a measure of probability used in hypothesis testing. Correlation has the following characteristics:

- the closer ρ is to zero, the weaker the correlation is;
- a positiveρ indicates a positive correlation, in which the values of the two variables tend to increase together;
- a negative ρ is indicative of a negative correlation, in which the value of one variable tends to increase when the other decreases;

One of the most common coefficients in statistics is the Pearson correlation, which can be defined as:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.12}$$

However, in the context of time series, we need to consider the study of Cross Correlations, which is a generalization of the correlation measure in that it takes into account the lag of one signal relative to the other. So, cross correlation is particularly important for assessing the causal relationship between two time series and keeping track of their similarity. It measures the similarity between two time series as a function of the time lag, allowing the uncovering of patterns and dependencies. If lag 0 is considered, the results will be equivalent to Pearson's correlation coefficient. Cross-correlation is useful because it can identify whether values in one time series are predictive of future values in another time series. In other words, it can tell us whether one time series is an anticipatory indicator for another time series. For example, it allows us to tell whether one year's value of a particular variable is predictive of that of another variable in subsequent years. Also in this case, the possible interval for the correlation coefficient of the time series is between -1 and +1. The characteristics are the same as the correlation that was listed above. The presence of a correlation is not sufficient to infer

the presence of a causal relationship. To infer a causal relationship between variables (in both directions), correlation cannot be used alone, as the conventional saying correlation does not imply causality implies.
In these cases, however, there is a definition of causality in time series that allows the sense of cause-and-effect to be determined and is based on the principle that the first series helps to predict the second. The earliest concept of causality for time series data was suggested by Granger[19].

2.2.7 Granger casuality test

The Granger causality test is a statistical test to determine whether one time series is useful for predicting another. It measures the extent to which the past values of one variable provide valuable information to predict the future behavior of the other variable. Granger causality helps to analyze potential relationships between variables, which helps to predict trends. If one can demonstrate through a series of t-tests and F-tests on delayed values of X (and with delayed values of Y included), then a time series X can be used to determine Granger causality Y.
The Granger causality test requires stationary data (which can be verified with the stationary tests presented previously). It is possible to transform non-stationary data into stationary data by differentiating them. A notion of causality based on how well past values of a time series y_t could predict future values of another series x_t. Underlying the test is a vector autoregressive model (VAR) of the Y endogenous time series[20]. Considering that we want to test that the time series X Granger-causes time series Y, two VAR models will be trained. The first, called the restricted model, assumes that Y depends linearly only on past values of itself, with linear coefficients y_i and a noise term time-dependent e_t :

$$Y_t = \gamma + \sum_{i=i}^{p} \gamma Y_{t-1} + e_t \tag{2.13}$$

The second model, on the other hand, also called the unrestricted model, assumes that Y depends linearly on the past values of X and Y, determined by coefficients α_i, β_i and a time-dependent noise term u_t:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-1} + \sum_{i=1}^{p} \beta_i X_{t-1} + u_t \tag{2.14}$$

The unformalized null hypothesis is that the second model adds no information, or provides a better model than Y, when comparing it with the first. However, this must be formalized into a testable hypothesis:

- Null Hypothesis (H0): time series X does not Granger-cause series Y
- Alternative Hypothesis (H1): the X time series Granger-causes the Y series.

If the p-value is less than a certain significance level (i.e., $p = 0.05$) then we can reject the null hypothesis and conclude that we have evidence sufficient to say that time series X Granger-causes time series Y. That is, the variance of the optimal prediction error of X is reduced by including the history of Y (informally, Y is causal of X if past values of Y improve the prediction of X). This characterization is clearly based on predictability and does not (directly) point to a causal effect of Y on X: Y improving the prediction of X does not mean Y causes X.

Chapter 3

Results

As introduced in the previous chapter the analysis of a time series allows to recognize the trend change points and seasonal change points.
In the results that will be presented, the BEAST algorithm has been used, to search only for trending change points since the minutes played by players under 21 are not seasonal.
The presence of trends in the data will then be verified through the use of the Mann-Kendall test, which will also determine the direction of the trend.
Thus, the analysis focuses on determining whether the percentages of minutes played by under 21 players affect the revenue obtained in the football market by Serie A teams and whether the percentages of minutes played by Italian under 21 players have any effects on the results of the Italian national team.
To compare the correlation and casuality between the minutes played by the Italian under 21 players and the results of the Italian national team, the cross correlation function and the Granger casualty test were used. The same methods were applied to the revenue earned by Italian teams and the minutes played by under 21 players. The book[21] from Derrick and Timothy, discusses the concept of cross-correlation in signal processing and statistics. It emphasizes the importance of applying cross-correlation analysis to stationary series. Shojaie et al.[22] provides a comprehensive review of Granger

causality, discussing its limitations and recent advances that address various shortcomings of previous approaches. The article emphasizes the importance of stationary time series for valid inference on Granger causality. Then, the two stationarity tests based on unit roots, Augmented Dickey-Fuller test (DF) and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) will be applied

3.1 Change point detection under 21 players

Applying the BEAST algorithm to the percentage of minutes played by under 21 football players from the 1967/1968 season to 2022/2023 produces interesting results shown in Figure 3.1.

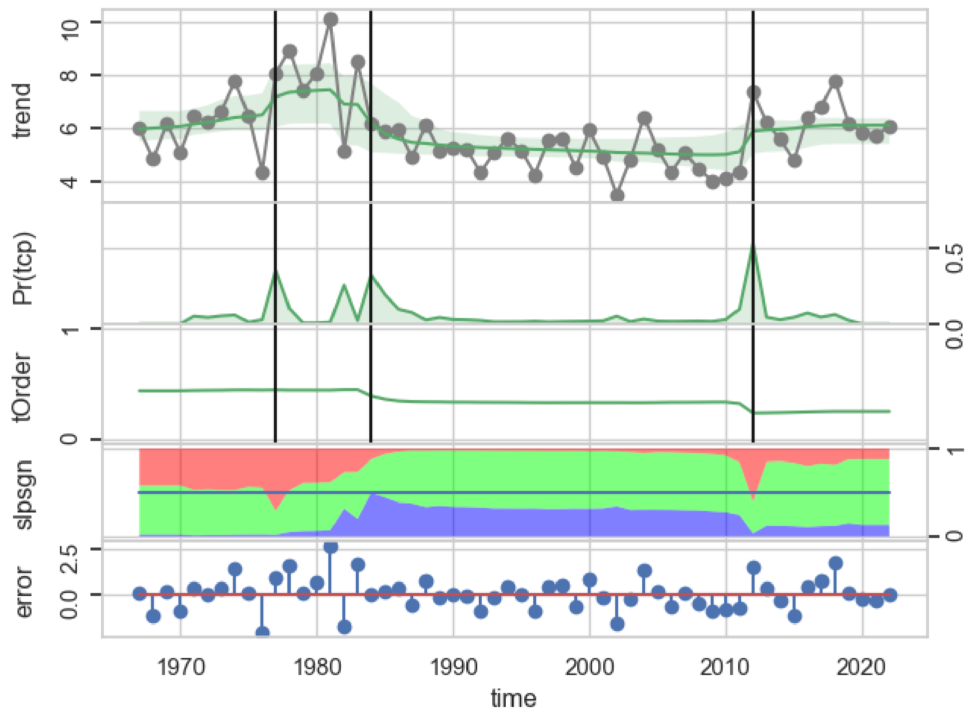

Figure 3.1: Bayesian change point detection in minutes played by under 21 players from 1967

The summary graph obtained is divided into five parts defined as follows:

- **trend** the trend and change points detected(vertical line),
- **Pr(tcp)** probability of trend change point occurrence overtime
- **tOrder** Time-varying polynomial order estimated to fit the trend(close to zero meaning a flat line rather than a sloped line)
- **slpsgn** probability of slope being positive(red part), zero (the green part), or negative(the blue part)
- **error**

As described in the previous chapter, the BEAST algorithm produces additional statistics when searching for change points. In fact, as can be seen in Table 3.1, the probability that a change point is present in a specific season is indicated.

tcp#	time (cp)	prob(cpPr)
1	1984	0.80096
2	2012	0.74767
3	1977	0.50188
4	1974	0.19608

Table 3.1: Probability of change point between 1967 and 2022

This probability is a result of the nature of the algorithm, which allows it to figure out in how many of the simulated MCM models the change point at time t is present. To understand how data distribution affects change point selection, it is important to analyze the standard deviation of our data within the mean in Figure 3.2.

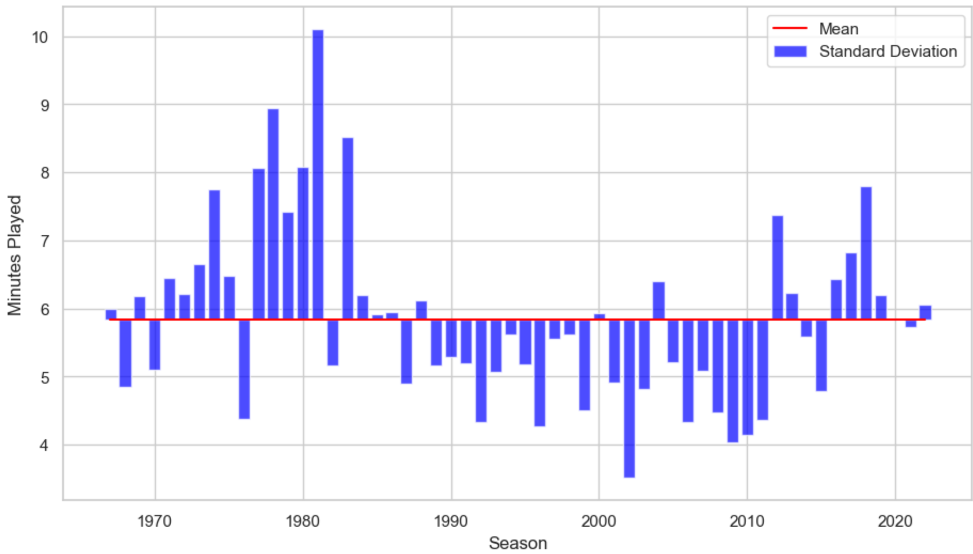

Figure 3.2: Mean and standard deviation

Season	μ	σ^2	95% Confidence interval
1974-1984	7.8074	1.3509	6.5093 - 9.1055
1984-2011	4.9593	0.4865	4.7469 - 5.3153
2012-2021	6.1007	0.9461	5.4827 - 6.7188

Table 3.2: Mean,variance and confidence interval of three main period

From the graph in Figure 3.1, we see three change points and a large period of stationarity.

- The first point of positive change is the result of the restrictive policies of Serie A towards foreign players in the early 70's. In fact, the policies that did not allow new foreign players in the teams took a few years to have the effect. The norm in fact, allowed foreign players in Serie A teams before the introduction of the standard to be able to make transfers between Serie A teams. Then the two change points in the percentage of minutes granted to players under 21 in the 1974 and 1977 seasons are the result of the protectionist policies implemented by the FIGC.

- The second point of change is related to the weakening of the restrictions on the number of foreign players allowed in Serie A. In fact, from the 1980-1981 season only one foreigner per team is granted, and then passed to 2 in 1982-1983. This rule has an impact and explains the change point in the 1984-1985 season. In fact, the slow and inexorable descent of the minutes granted to the under 21 players is visible, and is influenced by the new rule established by the FIGC that allowed three foreigners per team from 1987-1988.

- The long period of stationary starts from the early 1990s until 2011, and was not affected by Bosman's rule of 1996. Indeed, during these seasons the only change in the rules that could have generated changes in the minutes granted to under 21 is the European Court of Justice ruling on the Bosman case, which for the first time applies the rules on the free movement of workers within the European Union. This rule does not affect the minutes granted to under 21 players, but as will be analyzed later has an impact on the minutes granted to Italian under 21 players.

- The last chnage point in the 2012-2013 season shown in figure 3.1, is the result of the implementation of the new UEFA rules in the areas of economic sustainability and home grown players. In fact, Serie A from the 2011-2012 season introduced the home-grown rule created by UEFA, forcing the participating teams to have in the squad 4 players grown in their own team and 4 grown in their own league. The impact of Financial Fair Play has pushed teams to make a more sustainable transfer market, focusing more decisively on youth teams to complete their rosters.

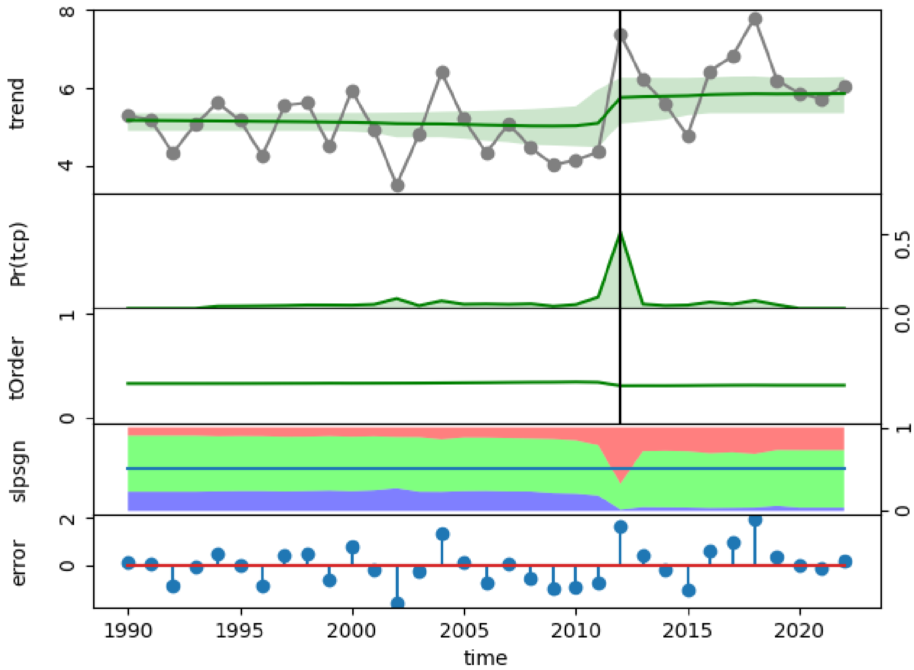

Figure 3.3: Bayesian change point detection in minutes played by under 21 players from 1990

3.2 Trend detection under 21 players

Due to the nature of the Mann-Kendall test, the tests must be applied at least starting two years before the changepoints. This explains why all tests are applied two years before the effective changepoint found by the BEAST algorithm. Applying the modified mann-kendall test to statistically evaluate the presence of the positive trend since the 1977 season and the negative trend from 1984.

Season	Z-value	P-value	Theil-Sen value	MK score	Decision
1972-1984	2.18979	0.02853	0.23459	41	Positive trend
1982-1992	-2.49119	0.01273	-0.18695	-33	Negative Trend

Table 3.3: Italian's under 21 result for the Mann-Kendall trend test

The Mann-Kendall test results offer a significant assessment of trends within the time series during the two specified periods. Both periods in question obtain a p value below the significance threshold. The test hypothesis of no trend in the data is thus rejected, thus confirming the presence of chanagepoints in the trend of minutes played by under21 football players. An increasing trend is then shown from 1972 to 1984, and a decreasing trend from 1984 to 1992.
The Mann-Kendall test can also be used to determine the absence of trends in the data, and verify a period of stationarity.

Season	Z-value	P-value	Theil-Sen value	MK score	Decision
1990-2011	-0.914720	0.36033	-0.02004	-18	No trend

Table 3.4: Italian's under 21 1990-2011 result for the Mann-Kendall trend test

As it is possible to see the long period of stationary from 1992 to 2011 does not show any kind of trend. To establish the stationary nature of the data were two of the most famous statistical tests: the Kwiatkowski-Phillips-Schmidt-Shin test and the Dickey-Fuller Increased test. The result is shown in the table, also indicating the p-values obtained from the tests.

Season	ADF test	*p-value*	KPSS test	*p-value*
1990-2011	-4.603778	0.00012	0.295077	0.1000

Table 3.5: Stationary test for Italian's under 21 1991-2011

ADF tests have a p-value of less than 0.05, so we can reject the Null hypothesis and conclude that the series is stationary. The null and alternate hypothesis for the KPSS test are opposite that of the ADF test. Then the KPSS test returns a p-value higher than the significance threshold of 0.05, and then the null hypothesis is confirmed, that the process has a steady trend. It is always better to apply both the tests, so that it can be ensured that the series is Truly stationary. Both tests show that the seasons from 1990 to 2011 are stationary and do not show trends in data distribution.

The last change point identified in the 2012-2013 season will be analyzed in more detail. In fact, the rules introduced first by UEFA and then by the FIGC were intended to increase opportunities for players raised in their leagues regardless of their nationality.

Season	Z-value	P-value	Theil-Sen value	MK score	Decision
2010-2023	2.5748	0.0100	0.1530	27	Positive trend

Table 3.6: Under 21 result for the Mann-Kendall trend test

The results of Mann-Kendall test confirm the increasing trend from season 2012-2013 for the minutes played by under 21 players, with a *p-value* of 0.0100 below the significance level. Therefore, the hypothesis *H0* is rejected, so there is a statistically significant increasing variation in minutes played by under 21 players in the Serie A league from the 2012-2013 season to 2022-2023.

3.3 Change point detection for Italian under 21 players

The search for change points in the percentage of minutes played by Italian under 21 players has produced results very similar to those obtained by analyzing all the minutes played by the under 21. In fact, until the Bosman law almost all the under 21 in Serie A were of Italian nationality. So to do a more detailed analysis it is interesting to analyze the data from 1990, a few years before the introduction of Bosman.

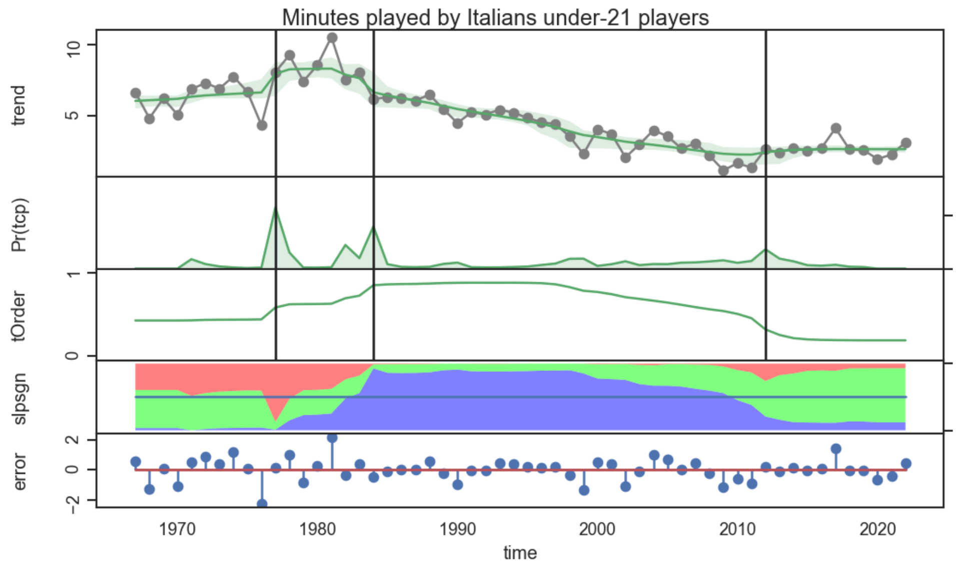

Figure 3.4: Bayesian change point detection in minutes played by italians under 21 players

The percentage of minutes played by Italian under 21 players in figure 3.5 since 1990 shows two change points. The first is in correspondence with the Bosman rule which, by liberalising the movement of footballers within Europe, gave the possibility of increasing the number of footballers in the youth teams. The other change point coincides with the home grown rule created by UEFA and applied by the FIGC from the 2011-2012 season.

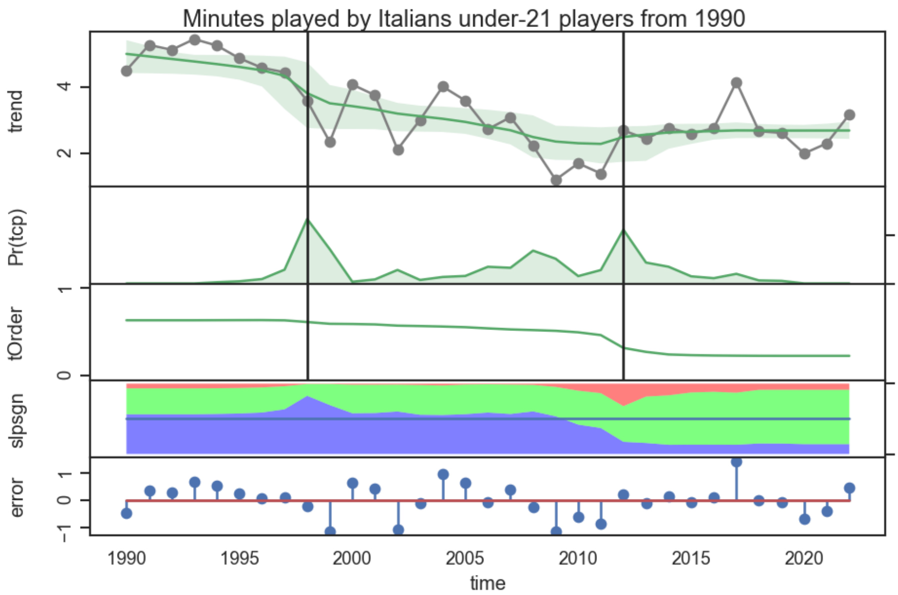

Figure 3.5: Bayesian change point detection in minutes played by italians under 21 players from 1990

Applying the modified Mann-Kendall test to statistically assess the negative trend since the 1996-97 season and the positive trend since the 2012-2013.

Season	Z-value	P-value	Theil-Sen value	MK score	Decision
1990-2003	-2.86741	0.00413	-0.21135	-48	Negative trend

Table 3.7: Result for italian players of Mann-Kendall trend test 1990-2003

The results obtained confirm the presence of a negative trend in conjunction with the Bosman law.
The home grown rule applied by Serie A shows a change point in the percentage of minutes allowed to Italian under 21 players, but it is not possible to establish the presence of a trend in the data as shown by the results obtained in table 3.4.

Season	Z-value	P-value	Theil-Sen value	MK score	Decision
2010-2023	1.781534	0.07482	0.08178	37	No trend

Table 3.8: Italian's under 21 result for the Mann-Kendall trend test

In fact, as you can see from table 3.8, the Mann-Kendall test returns a p-value of 0.75456. Failing to reach the threshold of significance of 5%, the null hypothesis of absence of a trend in data is confirmed.

3.4 Minutes played by Italians under 21 players and National Team

The minutes given to Italian under 21 players in Serie A can have an impact on the results achieved by the Italian national team and the national under 21 team. In fact, allowing more Italian players to play permanently in the top Serie A league can report effects on the results obtained by the national team on international stages. Thus, we assume that there is a correlation and Granger causality between the percentage of minutes given to Italian under 21 players and the results obtained by the Italian national team. The normalized score of Italy in the Fifa Ranking is used to evaluate the results of the major national team. The score of the Italian National under 21 team, is formed by the average score of matches played plus a factor of success in the European under 21 championship. The data analyzed refer to the loss that goes from 1993 to 2022, since the FifaRanking is a system of evaluation of the National teams established in 1993. The score of the Italian under 21 national team, was considered from the same year, in order to make a comparative analysis. As described above, the stationarity of the time series must be verified.

Time series	ADF test	*p-value*	KPSS test	*p-value*
Mins_played_italians	-3.3464	0.0129	0.4337	0.0626
Points_naz_u21	-5.2110	0.0000	0.3955	0.0791
Fifa_points	-3.0866	0.0276	0.1436	0.1000

Table 3.9: Stationary test for National Team and Minutes played by Italians

Table 3.9 shows the results of the stationarity tests for the three time series from the year 1993 to 2022:

- Minutes Played by Italians Under21 Players
- Points of the Italian National Under21 team
- Fifa Points of the Italian National team

3.4.1 Cross Correlation function

To carry out this analysis, the Pearson correlation indices and Cross Correlations described above were calculated.

Time Series	Pearson	Lag0	Lag1	Lag2
NazUnder21Result- Italians U21 Minutes	0.4265	0.4264	0.2931	0.2027
NazFifaPoints - Italians U21 Minutes	0.0261	0.0261	0.0176	0.0238

Table 3.10: Result of Cross Correlation for Italians U21 and National Team

The results shown in Table 3.10, allow us to analyze a positive Pearson correlation between the minutes played by Italian under 21 soccer players and the results obtained by the Italian under 21 national team. Using the cross-correlation function we notice a decrease in correlation at lag 1 and 2. This result, allows us to assert that the minutes played by Italian under 21 soccer players in the past, although in a small way have an influence on the results obtained by the Italian under 21 national team. Regarding the correlation between the FifaPoints obtained by the Italian national team and the minutes played by Italian under 21 football players, the results produced do not allow to affirm an effective correlation, because the obtained values are too low.

3.4.2 Granger causality

To determine whether there is a Granger causality between the minutes played by Italian under 21 players and the results of the Italian National Team are analyzed the results obtained by the Granger Causality test conducted.

H_0	Lag	Test Statistic	P-value
	1	12.8676	0.0049
Mins Italians U21 do not cause PointsNazU21	2	13.3443	0.0097
	3	11.9571	0.0353

Table 3.11: Granger Causality Results Points obtained by Italy national under-21 team and Minutes played by Italian U21 players

In the case of the hypothesis: minutes played by Italian under 21 players not cause the results of the Italian national football team, the null hypothesis is rejected given the p-value values below the threshold of significance of 5% for lag 1,2 and 3 shown in table 3.11.

H_0	Lag	Test Statistic	P-value
	1	1.0930	0.5789
Mins Italians U21 do not cause FifaPoints	2	1.2207	0.7480
	3	2.0921	0.7188

Table 3.12: Granger causality result FifaPoints and Minutes played by Italian U21 players

Table 3.10 shows the results of the Granger-Casuality Test to determine the casuality of the FifaPoints obtained by the Italian national team. The null hypothesis is accepted and therefore it is possible to affirm that the minutes played by Italian Under 21 players do not causes FifaPoints.

3.5 Minutes played by under 21 players and Revenue

The objective of the analysis of the correlation in question is to analyse a theme that emerged predominantly from the introduction of the Financial Fair Play[10]. That is to understand, if there is some kind of relationship between the revenues obtained by the teams in the league during the transfer sessions and the minutes granted to the under 21 players. In this analysis, only the gross revenues obtained from under-23 by Serie A clubs were considered. The time period analysed is again from 1993 to 2022. Data relating to the figures of transfers prior to this date are almost non-existent, since the value of the transfers on the transfermarkt site are entered manually after receiving the proposal of any user. For many past transfers the figures are not known, so it was decided to consider the data from 1993.

Time series	ADF test	*p-value*	KPSS test	*p-value*
Mins_played_under21	-3.3542	0.0126	0.3673	0.0912
revenue	-3.8396	0.0025	0.3944	0.0796

Table 3.13: Stationary test for revenue and Mins played under21

Table 3.13 shows the results of the stationarity tests for Revenue of Serie A teams and Minutes played by under21 players from the year 1996 to 2022.

3.5.1 Cross Correlation function

The results of the Pearson correlation coefficient and cross correlation are presented in Table 3.14

Time Series	Pearson	Lag0	Lag1	Lag2
Revenue - Minutes Played by u21 players	0.324	0.324	0.4178	0.1756

Table 3.14: Result of Cross Correlation for U21 players and Revenue

The results show a positive correlation between the minutes played by under 21 players and the revenues obtained by Serie A teams. The values obtained by cross correlation show that past values have a greater influence, in fact at lag 1 the coefficient of correlation increases. An easily explainable result, from the typical sale phenomenon of a promising young player at the end of a season characterized by high percentages of minutes played.

3.5.2 Granger causality

Referring to the objective of the analysis, it is expected that the results of the Granger-Causality test conducted would lead to the rejection of the null hypothesis, with p-values below of the significance level set at 0.05, thus demonstrating that minutes played by under 21 Granger-caused football players would be the Revenues of the Serie A teams.

H_0	Lag	Test Statistic	P-value
U21 Minutes do not cause Revenue	1	10.3037	0.0057
	2	13.6449	0.0034
	3	21.8637	0.0002

Table 3.15: Granger causality result Under21 Minutes and Revenue

The values obtained by the Granger-Casuality test shown in Table 3.15, reject the null hypothesis, thus confirming that the minutes played by under 21 Granger-cause the Revenues of Serie A teams in the transfer market.

Chapter 4

Discussion

The study presented aims to analyze the distribution of playing time for under 21 players by analyzing trends in the percentage of minutes given to under 21 players, following the various changes in the Serie A regulations that have occurred over the seasons. Two correlation and casuality analyses were performed starting with the playing time of young players in Italy's top league. The percentage of minutes played by Italian under 21 soccer players was used to understand a possible relationship with the results obtained by the Italian national soccer team. On the other hand, minutes played by under 21 soccer players were used to analyze the revenue obtained in the football market by Italian teams. Starting from the research questions, the following results were obtained:

- the regulation changes have strongly influenced the minutes played by Italian under 21 players and under 21 players. A positive trend is present over the past decade, particularly in minutes played by under 21 players.

- the profits made in the soccer market sessions by Italian teams from the sale of under23 players are correlated and caused by the minutes granted to under21 players.

- the results of the Italian under 21 national team are related to and caused by the minutes played by Italian under 21 players. In contrast, the Italian senior national team has no correlation and casuality with the minutes played by Italian under21 footballers.

The positive trend in the percentage of minutes granted to under 21 players that began in the early 1970s is a simple consequence of restrictive rules on the number of foreign players allowed in Serie A teams. The relaxation of these restrictive policies on the number of foreigners has led to a decrease in the minutes granted to under 21 players since 1984. The positive trend observed in under 21 player minutes played in Serie A from the 2012-2013 season onwards can be attributed, in part, to the introduction of the Home-grown Player Rule [16] by UEFA in the 2008-2009 season. This regulation was implemented with the aim of fostering the development of local talent and nurturing a stronger connection between football clubs and their communities. This rule required having a minimum of eight home-grown players (four grown in the same club and four grown in their own league) on the 25-man team. The same rule was then adopted by the Serie A starting with the 2011-2012 season. In Serie A, the influence of this rule has likely contributed to an increased emphasis on growing and fielding young, domestically-trained players. As clubs adapted to meet the Homegrown Player Rule requirements, a positive impact on the playing time and development opportunities for under 21 players within the league became evident. Despite the fact that the rule did not force a minimum number of minutes to be given to home-grown players, it is possible to highlight outstanding results achieved through the Home Grown Rule.
Another UEFA rule that has certainly had an impact is Financial Fair Play: originally introduced in 2009, UEFA's concept of financial sustainability has helped to drastically reduce club losses. In order to decrease operating expenses, many clubs began to invest in their youth sectors, from which they drew to complete their rosters and in many cases earn big money through the sale of promising home-grown youngsters.[1]

An example of this policy is Atalanta BC, which in the last 10 years, thanks to the players of the youth team debuted in the first team, has obtained excellent revenues and established itself as one of the best teams during the Serie A seasons. By analyzing the minutes played by Italian under 21 soccer players, it was possible to show a descescent trend, leading up to the Bosman law, which substantially decreased opportunities for young Italian soccer players. The introduction of the Home-Grown Rule, did not report statistically significant changes in the percentage of minutes played by Italian under 21 soccer players. In fact, the rule was created to increase the opportunities granted to home-grown players regardless of their nationality. Following the need to increase revenues, introduced by FFP, it was investigated what kind of relationship exists between the minutes played by under 21 players and the total revenues obtained by Serie A teams in soccer market. The analyses carried out showed a correlation between the two phenomena, but more importantly a causality, being able to state that the minutes played by under 21 players granger-cause the total revenues of the Italian soccer market. To disprove the correlation between the Italian national team's fluctuating results and the minutes given to young Italian players, the analysis conducted yielded partial results. The results of the Italian national soccer team cannot be correlated and are not granger-caused by the amount of minutes given to Italian under 21 soccer players. This result, is easily explained by the average age of the senior national team, which according to the latest survey is 27.4 years. However, this value must be put into context in the landscape of national teams, which historically prefer to rely on experienced players rather than on young talented players. For this reason, the score of the under 21 Italian national team was also analyzed. From the tests conducted, minutes played by Italian under 21 players obtain a positive cross-correlation with the points gained by under 21 Italian national team. Furthermore, after applying the Granger-Casuality test, it is possible to state that minutes played by Italian under 21 footballers, granger cause the results of the Italian under 21 national team.

4.1 Conclusion

In conclusion it is possible to say that the cliché, according to which in Serie A, no opportunities are granted to under 21 players is denied by the growing trend present in the last ten years. Even if the current percentages of minutes played by under 21 players in Serie A are lower than those of the 70s and 80s, the effects produced by the Home Grown Rule and the Financial Fair Play rules over the last ten years are very encouraging. In addition, by making a comparison with the major European leagues, as evidenced by OPTA Analyst[23](one of the most famous sites for the analysis and modeling of football data) the Serie A grants more minutes to players under 21 than the Premier League. In this special ranking the Italian championship is in fourth place, not far from the Bundesliga and LaLiga.

The effects produced by the minutes played by under 21 players, are a correlation and casuality in the revenues of Serie A teams. While the minutes played by Italian under 21 players have a correlation and casuality only on the results of the Italian under 21 national team.

4.2 Future works

This study can be seen as a starting point, with the possibility of being extended in various ways. Expanding the analysis, with a comparative study with other major European leagues beyond the range of minutes granted to players under 21 can reveal further insights. Studying global development strategies, the effectiveness of youth academies and the integration of young talents in the first teams could contribute to a wider understanding of youth development in European football. Also understand whether different types of national regulations have an impact on the actual minutes granted to young players.
Integrate the data of the minutes played by under 21 in European competitions, to understand if there is a disparity between the opportunities granted

in the national championship compared to the European stages. Considering also the Uefa Ranking [24] indicator (every season points obtained by a country's teams in all European competitions are calculated and then divided by the number of participating teams) to determine the impact of the under 21 players.

In recent years the data collected for each game is becoming more and more voluminous. In fact, for each player are no longer recorded only any goals and assists but also the completed passes, defensive interventions, and a factor expected goal (can be calculated in various ways), in addition to several other information. All this data, if analysed correctly, can be used to determine which roles and characteristics the most promising under 21 players have. It is also possible to define if the various teams follow patterns in the selection of the under 21 players integrated in the first team.

Finally, as proposed by Leyhr et al. [25], it is possible to implement diagnostics to measure player motor performance (sprinting, agility, dribbling, ball control, shooting) in young soccer players in order to study future player success. A predictive analysis aimed at selecting promising young players when they are still in the adolescent stage in under-15 teams.

Bibliography

1. Footballbenchmark. *New Financial Fair Play Regulations* https://www.footballbenchmark.com/library/overview_of_the_new_uefa_financial_fair_play_regulations.

2. Roel Vaeyens, A. C. & Philippaerts, R. M. Evaluation of the "under-21 rule": Do young adult soccer players benefit? *Journal of Sports Sciences* **23.** PMID: 16194977, 1003–1012. eprint: https://doi.org/10.1080/02640410400023266. https://doi.org/10.1080/02640410400023266 (2005).

3. C-415/93, C. Judgment of the Court of 15 December 1995. - Union royale belge des sociétés de football association ASBL v Jean-Marc Bosman.

4. Marcén, M. The impact of the Bosman ruling on the market for native soccer players (2014).

5. Blanpain, R. & Inston., R. The Bosman Case: The End of the Transfer System? *Leuven: Peeters* (1996).

6. Telegraph, T. *Uefa say Bosman ruined football's values* https://www.telegraph.co.uk/sport/2369701/Uefa-say-Bosman-ruined-footballs-values.html.

7. Binder, J. J. & Findlay, M. The Effects of the Bosman Ruling on National and Club Teams in Europe. *Journal of Sports Economics* **13,** 107–129. https://doi.org/10.1177/1527002511400278 (2012).

8. FIGC. *Incentivazione e promozione calciatori locali Serie A*

9. Bullough, S., Moore, R., Goldsmith, S. & Edmondson, L. Player migration and opportunity: Examining the efficacy of the UEFA home-grown rule in six European football leagues. *International journal of Sports Science & Coaching* **11** (Sept. 2016).

10. Regoliosi, C. Shedding light on the profitability of Italian professional football clubs where a different business model is performing. *International Journal of Sport Management and Marketing* **18,** 130 (Jan. 2018).

11. FIFA. *FIFA Ranking* `https://www.fifa.com/fifa-world-ranking/men?dateId=id13442`.

12. UEFA. *Under-21 on Serie A*

13. Bank, T. W. *Consumer price index Italy* `https://data.worldbank.org/indicator/FP.CPI.TOTL?locations=IT`.

14. Brockwell, P. J. & Davis, R. A. *Introduction to time series and forecasting* (Springer, 2016).

15. Zhao, K. *et al.* Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment* **232,** 111181. ISSN: 0034-4257. `https://www.sciencedirect.com/science/article/pii/S0034425719301853` (2019).

16. UEFA. *Protection of young players* `https://www.uefa.com/insideuefa/protecting-the-game/protection-young-players/`.

17. Hamed, K. H. & Ramachandra Rao, A. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* **204,** 182–196. ISSN: 0022-1694. `https://www.sciencedirect.com/science/article/pii/S002216949700125X` (1998).

18. Dickey, D. & Fuller, W. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *JASA. Journal of the American Statistical Association* **74** (June 1979).

19. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society,* 424–438 (1969).

20. Lütkepohl, H. *New introduction to multiple time series analysis* (Springer Science & Business Media, 2005).

21. Derrick, T. in, 189–205 (Jan. 2004). ISBN: 9780736044677.

22. Shojaie, A. & Fox, E. B. *Granger Causality: A Review and Recent Advances* 2021. arXiv: 2105.02675.

23. Analyst, O. *How Much Faith Does Your Club Show in Young Players?*

24. UEFA. *Uefa Ranking*

25. Leyhr, D., Kelava, A., Raabe, J. & Honer, O. Longitudinal motor performance development in early adolescence and its relationship to adult success: An 8-year prospective study of highly talented soccer players. *PLoS One* **13.** doi:10.1371/journal.pone.0196324 (2018).

Ringraziamenti

Un sentito ringraziamento va al mio relatore, il professore Marco Roccetti, che mi ha seguito, con disponibilità, fin dalla scelta dell'argomento, permettendomi di appassionarmi ad un promettente ambito di ricerca.
Un profondo ringraziamento va ai miei genitori, che mi hanno sempre sostenuto e che da sempre mi hanno insegnato quanto sia importante l'educazione e il rispetto come principio fondamentale nella vita di una persona.
Un grande ringraziamento va a mia sorella e mio fratello, sempre presenti anche se a distanza.
Infine un ringraziamento speciale va a tutti gli amici conosciuti a Bologna nel corso di questi due anni, che senza accorgermene hanno trasformato una città all'apparenza molto diversa da me in una seconda casa.