Università degli Studi di Salerno

Dipartimento di Scienze Economiche e Statistiche

Dottorato di Ricerca in:

Economia e Politiche dei Mercati e delle Imprese

Economia del Settore Pubblico

XXXI Ciclo

Tesi di dottorato in

The Multifaceted Aspects of Inequality: Health and Labour Market Issues

Dottorando:

Francesco Principe

Mat. 8801000012

Tutor:

Prof. Vincenzo Carrieri

Coordinatore:

Prof. Sergio Destefanis

Anno accademico 2017/2018

Contents

OVERVIEW

Inequalities have been the subject of keen economic interest due to ethical and policy implications for the society. In fact, there is wide agreement about the fact that living in a more egalitarian society leads to better social outcomes. However, on the other side, since most of the inequality is generated within the labour market there are discording opinions about the possibility of public intervention in the contemporary market economies (Machin 1997, Piketty and Saez, 2006). To make this picture even more multifaceted, related to economic inequality there is overwhelming evidence on the socioeconomic gradient in health. Indeed, economists have deeply investigated the two-way relationships between income and health and tried to disentangle the main mechanisms of transmissions (Deaton 2003). However, most of the findings rely on descriptive studies or analysis "at the mean". With the aim of providing causal evidence and contributing to this long-standing strand of literature, this thesis analyses distributional aspect of income, focusing on the top tail of income distribution and in particular on those individuals who get high earnings in the labour market: the so-called working super-rich. Moreover, it investigates the socioeconomic gradient in health both in the top tails and in the rest of the distribution. The thesis consists of three chapters.

The first chapter focuses on the determinants that allow some individuals to receive extraordinary earnings in labour markets compared to their peers, i.e. CEO or the superstars of sport, music and cinema. Specifically, it empirically analyses the effects of performance, popularity and bargaining power on the earnings of the universe of football players of Italian Serie A. The reasons are multiple. First, exploiting the possibility of having a perfect match employer-employee. Second, a longitudinal dataset with detailed information about the football players of Italian Serie A. Lastly, the fact that in Italy football players and managers represent the 20% of top 500 earners. On a methodological point of view, unlike previous analyses that are essentially based on cross-sectional data, disposing of panel data allows to investigate the returns of performance, popularity and bargaining power while controlling for players and team unobserved heterogeneity. Moreover, I employ the Unconditional Quantile Regression approach developed by Firpo, Fortin and Lemieux (2009) to estimate the impact of a marginal change in the determinants of earnings on their entire distribution. This is relevant because players' earnings exhibit a large dispersion around the mean and investigating the role of the determinants is especially interesting at the top of the earnings distribution, where superstar effects should more clearly manifest. Main results show that all the aforementioned factors significantly affect the players' earnings. However, the analysis "beyond the mean" reveals that the role played by popularity increases at the top of earnings distribution being the main determinant of the "superstars". These results challenge the interpretations of extraordinary earnings based only on very talented workers who "win and take all".

As mentioned above, labour market seems a fertile ground for the escalation of contemporary society's extreme inequalities. Thus, understanding how the "working super-rich" respond to health shocks can be a key element of these dynamics. The second chapter of this thesis aims at providing evidence of the relationship between health shocks and labour market outcomes, focusing on those in the top tail of earnings distribution. Therefore, the dataset presented above has been enriched with data about the nature and the incidence of the injuries suffered by the universe of Serie A football players. In particular, in this chapter, I exploit traumatic injuries as exogenous variation in professional football players' health to provide estimates of the causal impact of a health shock on two main labour market outcomes: the annual net wages and the probability of changing the terms of the contract between the employer (the club) and the employee (the player).

The empirical approach employs panel fixed effects models combined with an IV strategy, which uses the average number of yellow cards received by the team as an instrument. Main results show that working super-rich are not immune to economics consequences of health shocks. In fact, injuries reduce the net wage in the following season by around 12%. This result is mainly driven by precautionary reasons due to the club's concern about depreciation in the player's human capital rather than by a direct effect of the shock on the player's productivity.

Thus, the second chapter contributes to the existing literature by providing original findings that can be summarised as follows. Firstly, it provides causal evidence of the relationship between health shocks and labour market outcomes. Secondly, it focuses on the consequences of health shocks for those on top incomes, for whom there is scant evidence so far. Lastly, it allows for a deeper analysis of the main mechanisms of the health shock, disentangling the effect mediated through the player's performance and the one generated by human capital depreciation, inducing the club, to offer a lower wage for precautionary reasons.

The third and last chapter carry on the emphasis on inequality and its multifaceted aspects by moving the focus from the top tail to the whole population. In fact, it investigates how people heterogeneously respond to public education policies delivering health information by using as case study the 2015 WHO warning about the carcinogenic effects of red meat consumption. Importantly, it exploits high frequency data, i.e. monthly data, about Italian households' expenditures to identify the effect of the warning in the long vs. short run, for which there is no previous evidence in the literature and to document the response of households differing due to educational levels and health awareness. In order to identify such effect, I employ a Difference-in-Difference model which exploits the strong seasonality patterns in meat consumption in Italy, mainly associated with culinary traditions in occasion of catholic holidays celebrations. The results show a general misinterpretation of the warning and a short-term fall in red meat consumption by around 5%. However, a long-lasting and consistent shift in red meat consumption is documented only among households with higher educational levels and health awareness. These findings highlight a brand-new driver of health-education gradient and a potential source of health inequalities. Finally, they have strong policy implications about the successful design of public health-information policies, suggesting that they should be designed in a way that expose the individual to a constant flow of information.

References

Deaton, A. (2003). Health, inequality, and economic development. *Journal of economic literature*, *41*(1), 113-158.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953-973.

Machin, S. (1997). The decline of labour market institutions and the rise in wage inequality in Britain. *European Economic Review*, *41*(3-5), 647-657.

Piketty, T., & Saez, E. (2006). The evolution of top incomes: a historical and international perspective. *American economic review*, *96*(2), 200-205.

CHAPTER I

What makes you ‘super-rich’? New evidence from an analysis of football players’ wages

Abstract

This article investigates the influence of performance, popularity and bargaining power on “super-earnings” using a unique panel dataset of Italian football players built on various sources of data. Using OLS, Panel and Unconditional Quantile regression techniques, I find that detailed measures of these factors are all significantly associated with higher wages. Popularity dominates all the other factors at the right tail of earnings distribution and the agent’s power contributes mostly to allocate players in richer teams. These new findings challenge the interpretations of super-earnings based only on very talented workers who “win and take all”.

An extract from this chapter is published as: Carrieri, V., Principe, F., & Raitano, M. (2018). What makes you ‘super-rich’? New evidence from an analysis of football players’ wages. *Oxford Economic Papers*, *70*(4), 950-973.

1. Introduction

Since the middle of the 1970s, the share of gross personal income held by the top 1% or 0.1% of the population sharply increased in some developed countries, especially in the US and the UK, bringing the incomes of the top groups back to the levels they achieved at the beginning of the twentieth century (Atkinson et al. 2011). Furthermore, observing the composition of top incomes reveals a striking novelty. In fact, while in the past, the large majority of individuals belonging to the richest segment of the population included rentiers or entrepreneurs (Alvaredo et al. 2013), there has been a considerable increase in the number of the "working super-rich" accessing the top income bracket in recent decades. For instance, for the richest 1% of the population, earnings accounted for the 46.4% of the total in 1980 in Italy while they accounted for 70.9% of the total in 2008 and, similarly, among the 0.1% richest segment of the population, the share of earnings rose from 29.5% to 66.2% in the same period (Franzini et al. 2016). Hence, in contemporary economies, the labour market seems to be where extreme inequalities grow.

The economic literature has addressed the phenomenon of the working super-rich by offering explanations based essentially on individual talent or popularity and, more recently, on bargaining power (see Section 2 for more details). The seminal works of Rosen (1981) and Adler (1985) noted the role of individual talent and popularity, respectively, and they argue that super-earnings emerge from fierce competition among the best performers in sectors where technology magnifies the earnings of the winner by allowing joint consumption. On the other side, more recent interpretations of the extraordinary rising salaries of top managers in large companies refer to power exerted by them on shareholders in contexts characterised by asymmetric information (Bebchuck and Fried 2003). Moreover, some other contributions have also noted that the superstar status might not be always related to abilities crucial for the specific type of performance in which individuals are involved (i.e., Franzini et al., 2016). Rather, it might be assimilated to a rent because notoriety and conformist behaviours by consumers assign to some superstars the possibility of extracting rents unrelated to their talent or their effective current productivity.

Perhaps due to the difficulty of finding good proxies, no empirical studies – to my knowledge – have inquired on the joint influence of talent, popularity and bargaining power on "super-earnings". The primary goal of this chapter is to fill this gap. I focus on a category of working super-rich – football players – who represent a consistent share of the universe of super-rich along with some heterogeneous professional categories, e.g., business lawyers, investment bankers, top managers working for large corporations, etc. (Atkinson et al. 2011). As noted by Kahn (2000), professional sport offers a unique opportunity for innovative labour market research[1], because several indicators about a player's characteristics and performance are widely available and salaries are regularly published by the dedicated press.

Moreover, the influence of talent, popularity and bargaining power on super-earnings can be easily discerned in professional team sports because teams also obtain returns by hiring players endowed by these characteristics. Indeed, team's owners are willing to pay talented players in order to increase both revenues from TV rights, tickets and merchandising and – not least – to enhance team performance. They are also willing to pay famous players since they can exploit their popularity by selling more tickets or through merchandising. In contrast, both talented players and less talented but "famous" players might be able to bargain a higher salary when a club's owner exploits the threat to choose a different team. The influence of these factors on earnings might actually have been reinforced by the technological and institutional changes that occurred in Europe in recent years. On one hand, pay-TV technology and the internet have allowed teams to be watched by a global audience and contributed to redistributing the largest share of the revenues towards the most popular teams and the most talented and famous players (Boeri and Severgnini 2012). On the other hand, institutional changes, such as the diffusion of free agency – i.e., the eligibility for a player to sign with any club even when under contract to a specific team[2] – and the Bosman ruling – which liberalised players' markets within the European Union by removing transfer fees when a player wanted to change clubs when the contract had expired – allowed players to strengthen their bargaining power, also relying upon professional agents in order to negotiate better deals with team's owners (Blair 2007, Mason 2012).

We use a longitudinal dataset on football players[3] in the Italian Premier League (*Serie A*) built by merging information from various sources of data about players' characteristics, performance and wages. The Italian Premier League (*Serie A*) is one of the top five most followed football leagues in Europe, and football players represent an important share of the super-rich, i.e., in 2003, they constituted approximately 1/5 of the top 0.01% earners in Italy (Franzini et al. 2016).

Additionally, due to the peculiarity of my dataset, my analysis allows us to make a number of contributions to the existing literature (reviewed in Section 2). First, unlike previous analyses that are essentially based on cross-sectional data, I can dispose of a panel dataset that allows us to investigate the returns of performance, popularity and bargaining power while controlling for players and team unobserved

[1] Besides, a number of studies use sport data to analyse various economic issues. Recent examples include Gallo et al. (2013) and Price and Wolfers (2010) to analyse discrimination in the labour market and Dickson et al. (2016) to analyse determinants of domestic violence.

[2] Zimbalist (1992) finds that players who are not eligible to become free agents suffer a higher monopsonistic exploitation (about 38%) by teams than their eligible colleagues (18%).

[3] Throughout this paper, I use the words "football" and "soccer".

heterogeneity. This approach permits us to establish a more causal link between determinants and earnings. Second, I use the Unconditional Quantile Regression approach developed by Firpo, Fortin and Lemieux (2009) to estimate the impact of a marginal change in the determinants of earnings on their entire distribution. This is relevant because players' earnings exhibit a large dispersion around the mean and investigating the role of the determinants is especially interesting at the top of the earnings distribution, where superstar effects should more clearly manifest. Third, I use several performance measures (goals, assists and average grades obtained during a season) as a proxy of talent, the number of yearly Google search queries made for each player as a proxy of popularity and the total market value of players who are represented by the same agent as a proxy of the bargaining power. In particular, unlike the previous literature, which usually makes use only of goals and assists as measures of performance, this allows us to properly evaluate the performance of all team members, including those who are not directly involved in goals or assists, such as midfielders or defenders. Moreover, the use of measures of power is new in the literature, and its role on determining player's earnings has been unexplored so far.

We find that all three aforementioned factors – i.e., performance, popularity and bargaining power – significantly affect players' earnings. These results are driven by both a pure compositional effect (i.e., the allocation of players endowed by a higher talent, popularity or power in better teams) and a pure direct effect, as the impact of these determinants on earnings is largely significant when players and team unobserved heterogeneity is taken into account. Moreover, analysis 'beyond the mean' reveals that especially the role played by popularity increases at the top of earnings distribution, among those who, according to the terminology developed by Rosen (1981), can be named as "superstars". These results challenge the interpretations of extraordinary earnings based only on very talented workers who "win and take all".

The rest of the chapter is structured as follows. In Section 2, I discuss the insights from the economic literature about possible explanations of super-earnings, briefly reviewing the contribution of sports economics to this literature. Section 3 presents my data and the main variables. Section 4 discusses the empirical methodology. In Section 5, I present the results of my empirical analysis. The last section summarises and concludes.

2. Determinants of super-earnings: insights from the literature

The evidence that a few individuals in selected professions – e.g., athletes, singers, artists, writers, CEOs, and lawyers – can enjoy huge salaries has been named in the economic literature as the "superstar" phenomenon (Rosen 1981). The economic literature has addressed this issue by offering few, and not always recent, explanations that can be classified according to the role they assign to the individual's talent (Rosen 1981), popularity (Adler 1985) or power (Bebchuck and Fried 2003). In this section, I offer a quick review of these contributions and discuss the main empirical findings of the literature trying to test these theories.

The seminal theory of superstar formation was provided by Rosen (1981), who showed how a small difference in individual productivity/talent can be magnified into huge differences in earnings, focusing on three main assumptions. First, consumers are able to identify who are the best performers. Second, they prefer to be served by "the best", i.e., there is imperfect substitution among performers. Third, technology allows for joint consumption – i.e., there is no rivalry among consumers – and better performers can draw large audiences, for instance, in football stadiums, or via TV or selling their books or albums worldwide, with a cost of production largely independent of the size of the audience. Krueger (2005), for instance, shows that these are salient features of the market for rock concerts and among the leading explanations for the high prices and revenues of music shows by superstar musicians. According to Rosen (1981), super-earnings depend exclusively on talent: in sectors such as professional sports markets, show business and many entertainment services (Rosen and Sanderson 2001), the most talented individual wins fierce competition and, independent of the size of the difference in productivity with the losers (that can also be very small), the winner takes most of the pie (as in the "winner takes all" markets discussed by Frank and Cook, 1995).

A second theory was proposed by Adler (1985), who argued that superstars might emerge among equally talented performers due to the positive network externalities of popularity. While, on the supply side, Rosen (1981) and Adler (1985) agree on the necessity of technologies allowing large economies of scale, Rosen (1981) considers talent to be observable without cost by all consumers, while Adler (1985) notes that talent is a hidden characteristic that has to be discovered through personal and interpersonal learning processes. Actually, the appreciation of a particular performer (e.g., football player, singer, or artist) grows with the knowledge consumers have acquired about him through conversations with other people. In fact, as a performer's popularity increases, it becomes easier to find other fans, because, due to searching costs, consumers are better off patronising the most popular star as long as others are not perceived as clearly superior. According to Adler (1985), luck determines who among equally talented performers will snowball into a star, but talent is an essential prerequisite to becoming a superstar. However, Adler (2006) states that the likelihood of becoming a superstar could also be affected by the investments performers make in their popularity, through advertising or appearing on talk shows, in tabloids, in magazines and on social networks.

Similarly, Franzini et al. (2016) recently argued that superstar status might not always be related to abilities crucial for the specific type of performance in which individuals are involved and might be assimilated to a rent. Especially for sport and show business stars, very high earnings can be generated by providing services through activities in which one does not necessarily excel. The well-known phenomenon of celebrity endorsement, indeed, guarantees very high revenues from advertising to a star. Furthermore, apart from mechanisms highlighted by Adler (1985), when information on individual abilities (e.g., of athletes, singers, managers and professionals) are largely imperfect, notoriety and conformist behaviours by consumers assign to some superstars the possibility of extracting rents unrelated to their talent or their effective current productivity. For instance, this may occur when, due to proven success in the past, popularity fails to provide information on current abilities, which can fade more quickly than fame[4]. Moreover, additional forms of rents can be gained by some superstars, influencing preferences and choices of consumers through advertising or exploiting popularity offered by the media. Thus, less talented individuals might also acquire popularity, challenging the view that superstars always emerge among the most talented.

As a third explanation, extraordinary earnings might be associated with the bargaining power exerted by superstars. For instance, managers and CEOs in large companies might be able to fix their own remuneration independent of productivity, exploiting asymmetric information with respect to shareholders (Bebchuck and Fried 2003, Bivens and Mishel 2013). Furthermore, players in team sports can achieve wage increases by threatening the club owner with acceptance of a better deal from another team (Blair 2012).

A number of empirical contributions have tried to investigate the determinants of super-earnings, and, with few exceptions, mostly rely on sport statistics, as mentioned in the introduction (see, e.g., Frick 2007 and Deutscher and Buschemann 2016 for a survey)[5]. However, most studies focusing on sport economics issues investigate players' characteristics that are associated with higher wages (e.g., the player's position, the footedness, the age and the experience in the League, or performance measures, such as tackles, assists or goals in soccer or points scored or rebounds in basketball) and are based on cross-sectional data and use OLS Mincerian (Mincer 1974) regressions or quantile regressions (Koenker and Bassett 1978).

Fewer studies explicitly test theoretical predictions, i.e., try to investigate the role of performance and popularity in super-earnings. Lucifora and Simmons (2003), among others, found evidence to support Rosen's explanation of superstars in the Italian soccer league, as they find that talent, measured by performance indicators, exercises significant influence on the skewness of the earnings distribution. More recently, other authors have studied the role of either talent or popularity in shaping players' earnings in sports, finding controversial results. Treme and Allen (2009) focused on drafting of rosters in the US National Basketball Association (NBA) and found a significant effect of both performance before being drafted and the media exposure received by players on entry earnings. Franck and Nuesch (2012) found that both talent- and nonperformance-related popularity increase the market value of soccer stars, especially at the top of the distribution; Lehmann and Schulze (2008) showed, instead, that neither performance nor popularity explain salaries of soccer superstars in the top quantiles[6].

4 Mullin and Dunn (2002) described the star's popularity of a baseball player as an intangible characteristic coming mainly from reputation based on past performance, which attracts fans who pay to see these stars, even when their playing performance is no better than mediocre.

5 A significant exception is represented, for instance, by Krueger (2005) who analyses the market for superstar musicians and finds a significantly positive association between star quality measured by the number of millimeters of print columns devoted to each artist in The Rolling Stone Encyclopedia of Rock & Roll and the prices and revenues of the music shows.

6 Following a different empirical strategy, Brandes et al. (2008) compared talent- and popularity-based explanations; more specifically, they compared star attraction of national superstars and of so-called "local heroes" in the German soccer league

To the best of my knowledge, no studies have tested the influence of an agent's bargaining power on earnings and none of the studies reviewed in this section has jointly analysed the role of performance (as a proxy of talent), popularity and bargaining power on superstar earnings. As mentioned in the introduction, the main goal of this chapter is to fill this gap.

3. Data and main variables

Our empirical analysis is based on an original dataset recording earnings, performance and other characteristics of football players of the Italian Premier League (*Serie A*). The data come from various sources and record longitudinal information for players, teams and agents[7]. I considered 469 players who appeared in *Serie A* in the 2013-2014 season as the starting sample – excluding goalkeepers, thus following a common approach in the sport economics literature (e.g., Lucifora and Simmons 2003) – and followed these players from the 2010-2011 season to the 2014-2015 season. This leads to a panel dataset composed of 1,586 observations. The panel is unbalanced because, due to the system of clubs' promotion and relegation between *Serie A* and *Serie B* and players' transfers across national and international leagues, there is a large turnover of players in the league.[8]

We built the dataset in order to observe all factors that might influence players' earnings, controlling for club and player characteristics: i.e., talent, proxied by measures of performance, popularity and bargaining power.

Data on players' yearly salaries – recorded net of taxes and excluding possible bonuses – are taken from the annual report published at the beginning of each season by the most-read Italian sport newspaper, *La Gazzetta dello Sport*. Data about players' characteristics (e.g., age, position in the pitch, and international caps) and performances (i.e., goals and assists – decisive passes leading to a goal – during a season) are extracted from the websites *transfermarkt.com* and *soccerways.com*. As a further performance variable, Ialso included the average seasonal grade assigned to players in each played match by the three most popular Italian sport newspapers – *La Gazzetta dello sport*, *Il corriere dello sport* and *Tuttosport* – where the grade varies between 1 (very poor performance) and 10 (excellent), even if, in most cases, journalists use a range

– defined as the most valued players of teams in which no national superstars play – and find that superstars attract fans by outstanding field performances, whereas local heroes facilitate fan support by mere popularity.

[7]Ialso collected information on teams' economic performance, recorded in balance sheets, annually approved by the clubs' boards of directors and published on the official websites. However, I prefer to include team fixed effects rather than specific values for team variables in my baseline estimates, as these variables are rather time invariant. Results, available upon request by the authors, show that my main findings do not change at all ifIreplace team fixed effects with values for a set of team variables.

[8] The Italian Premier league is composed of 20 teams, and at the end of the season, the last three clubs in the table are relegated and substituted by the first three clubs in the second division (*Serie B*).

between 4 and 8.[9] As mentioned in the introduction, this process allows us to measure the performance of players not directly involved in goals and assists, such as midfielders and defenders.

Concerning popularity, I use as a proxy the number of Google search queries made each year for each player.[10] As a proxy of players' bargaining power, I rely on information on the total market value of players who are represented by the same agent (provided by the website *transfermrkt.com*), assuming that an agent with a richer portfolio is better able to bargain a good deal with the club's owner. According to Italian rules, football players can be represented by professional agents enrolled in a specific register or by a close relative. The relationship between the player and the agent is generally very stable over the time. Actually, an agent generally follows the player's interest over his entire career, with changes in agents being quite rare, at least in relation to my period of observation (five seasons). Therefore, the proxy of agent bargaining power is, in fact, time invariant in my dataset.[11]

The contract between the team and the player ensures the right for the team to enjoy the sports performance of the player and to involve him in public events related to the sponsors and the club's image. On the other hand, the player is generally paid with a fixed salary and, occasionally, a variable that depends on the team and individual results during the football season. In this chapter, I consider only the fixed part of the salary.

Contracts between the team and the player can last for 5 years at most. Unfortunately, I do not have information about contract length in my dataset (neither the original nor the residual duration). However, evidence suggests that contracts in the Italian *Serie A* usually last for 5 years, with the exception of young and old players, that are often characterized by shorter-term arrangements (Carmichael et al. 2012)[12]. Contracts can be renegotiated even before the expiry date (often to contrast claims by other teams) and

---

9 Other recent papers in the sport economics literature use a player's grade as a proxy of performance, e.g., Bryson et al. (2012), Buraimo et al. (2015) and Deutscher and Buschemann (2016).

10 To have a coherent value for each player, the data have been collected the same day for each one, typing "name-surname-team". I also collected data about the number of followers each player has on Twitter, butIdid not rely on this measure because only the most popular players are on Twitter. The few studies that have investigated the link between football players' popularity and earnings use proxies based on players' quotations in newspapers (Brandes et al. 2008, Lehmann and Schulze 2008) or press publicity (Franck and Nuesch 2012).

11 Unfortunately, available statistics on professional agents (i.e., on the site transfermarket.com used for my analysis) do not include any information on the length of the contract as it is actually a part of the private contract between the player and the agent. However, I do not observe any change in the agent-player relationship in my data and this might be explained by a number of reasons. First, because my data refer to the period before April 2015, when, for the first time, a maximum length of two years has been set by the Italian Football Federation (FIGC) to the contract between the player and the agent with the aim of introducing more competition in the market for agents. The introduction of a maximum length of 2 years suggests a length well above this threshold in the period to which my analysis refers to (i.e. before 2015). Secondly, contracts between agents and players both before and after 2015 include the payment of a fee in case of anticipate resolution without a legitimate reason and this obviously creates a cost on both parts to the dissolution of the contract. Last but not least, the average observational period for each player is in fact of 3 years in my data and this is due to the system of relegation/promotion and to some cases of player's transfer to other leagues (especially as concerns foreign players).

12 Consistently with Carmichael et al. (2012), the inclusion of a careful control for the age of the player (age and age squared variables) in all my estimates allows us to partially take into account the effect of contract length on earnings. However, a precise estimate of the relationship between contract length and players' earnings, as that performed by Buraimo et al (2015) for the German Bundesliga, cannot be carried out through my dataset.

in some cases automatic earnings increases are established by the contract according, e.g., to the number of games played during a season. When looking at players' earnings changes overtime – a poor proxy of the contract length as explained above – my data show that approximately 63% of players receive a wage different from that of the previous season (56% if I exclude those moving to another team).

In addition to the proxies of individual talent, popularity and bargaining power, I also included in my dataset several variables that are used as controls in my estimations (see Sections 4 and 5) and that are presented, along with some summary statistics of all variables, in the next section.

3.1 Descriptive statistics

The full list of variables included in my dataset along with mean values and standard deviations are presented in Table 1. On average, the net annual earnings of football players in the Italian Premier league amount to approximately 875,000 Euros, but the standard deviation is very high (912,000 Euros). In addition, proxies of popularity and bargaining power are characterised by a very high standard deviation that is much higher than the mean of the variable (Table 1). On average, players in my sample score 1.93 goals per year (standard deviation is 3.50) and 1.28 assists (standard deviation is 2.04), while the mean grade assigned by journalists to players' performances in each match is 5.77 (standard deviation is 0.41).

[Table 1 approximately here]

The large dispersion in players' annual net wages clearly emerges when values of percentiles of earnings distribution and ratios among percentiles are shown (Table 2). Even if almost the whole body of football players earns a very high wage (the 10^{th} percentile earns 200,000 Euros per year), a group of "superstars" clearly emerges: the top 10% earn at least 2 million Euros per year and the top 5% earn at least 5 times more than the median earner (Table 2). Therefore, the Gini index within the group of football players in my sample is very high: it is only slightly below 0.50, while, for comparison, one may consider that the Gini of net annual earnings from dependent employment was in Italy 0.298 in 2014 according to EU-SILC 2015 data (0.239 if only full-time employees working for the whole year are considered).[13]

[Table 2 approximately here]

[13] However, my estimates of the Gini index are in line with results found for other soccer leagues. For instance, Franck and Nuesch (2006) find a Gini of 0.56 for the German Bundesliga in the 2004-2005 season (but referring to market values that include also bonuses and potential transfer fees and are then likely to be more dispersed than wages), while Send (2016) finds a Gini of 0.51 considering the earnings of the German Bundesliga players in the 2014-2015 season. A higher inequality emerges instead in the US Major Soccer League: Reilly and Witt find a Gini of 0.628 in 2007 (0.569 excluding from the computation Beckham who was the most paid player at that time).

Figure 1 shows the non-parametric estimate of the overall salary distribution for the pivotal 2013-2014 season. A very asymmetric distribution emerges, with a long right tail, which indicates the presence of a restricted number of players who earn very high salaries compared to the rest of the distribution. As mentioned, this is consistent with the "superstar" phenomenon discussed in Section 2.

[Figure 1 approximately here]

Apart from players, a large heterogeneity also characterises the *Serie A* teams. As stated in the following sections, the achievement of higher salaries by talented, popular or powerful players can be mediated by purchases by the richest teams, which can afford higher wage bills. Indeed, total wage bills hugely differ among clubs who participate in the Italian football Premier League (Figure 2) and because economic performances by teams and revenues from game tickets and TV rights are very unequally distributed across teams (Table 3).

[Table 3 approximately here]

4. Empirical strategy

Our empirical analysis relies on two main estimation techniques. I first assess the impact of main determinants of earnings 'at the mean' of the earnings distribution using both pooled OLS and panel estimators. Second, It est the impact of the determinants along the entire earnings distribution using unconditional quantile regressions.

4.1 OLS and panel analysis

Consistent with the empirical literature on athlete earnings, I estimate the following augmented Mincerian equation:

$$\boldsymbol{Ln(W_{it}) = \beta_0 + \beta_1 Perf_{it-1} + \beta_2 Pop_{it-1} + \beta_3 Power_{it-1} + \gamma X_{it-1} + \delta Season_t + \varepsilon_{it}} \quad \mathbf{(1)}$$

where the dependent variable is the log of yearly net player's earnings in season t, $\boldsymbol{Perf_{it-1}}$ is the vectors of proxy variables for players' performance in the previous season (i.e., goals, assists and average grades during the season), and $\boldsymbol{Pop_{it-1}}$ and $\boldsymbol{Power_{it-1}}$, respectively, refer to proxies of popularity and bargaining power, measured before the season starts. $\boldsymbol{X_{it-1}}$ represents a set of several lagged time-varying and time-invariant player's characteristics that I include as controls in my estimates, namely, age and age squared, the number of caps with the national team and with the under 21 national team during the previous season and on the whole during one's career, the number of minutes played during the previous season, dummies for the position in the pitch (distinguishing defenders, midfielders and forwards), dummies for citizenship (distinguishing Italian, EU and extra-EU players), a dummy for players who are captains of their teams and season dummies. All the regressors are in lagged values in order to rule out potential reverse causality issues.

As an additional model, I also estimate equation (1), adding team fixed effects to covariates (i.e., dummies for teams in season *t*), in order to capture possible heterogeneity in earnings related to the club's characteristics (e.g., prestige, wealth).

$$\boldsymbol{Ln(W_{it}) = \beta_0 + \beta_1 Perf_{it-1} + \beta_2 Pop_{it-1} + \beta_3 Power_{it-1} + \gamma X_{it-1} + \delta Season_t + \vartheta Team_t + \varepsilon_{it}\ (2)}$$

Equation (2) is useful to observe whether performance, popularity and bargaining power exert a "direct" influence on earnings or whether the influence of these three factors is merely compositional, or "indirect" – i.e., it is mediated by the likelihood of more talented, popular and powerful payers to belong to a richer team. For instance, popularity or an agent's power could allow players to achieve higher earnings, allowing them to be purchased by a better team, without then exerting a further "direct" effect within teams. Henceforth, estimates with team fixed effects can be interpreted as the estimate of a "within team" effect on earnings due to performance, popularity and power.

In order to eliminate from my estimates the possible confounding factor due to the correlation between performances and the other two determinants of super-earnings (i.e., performance-related media coverage and/or the capacity of most talented players to be represented by the most powerful agents), as suggested by Franck and Nuesch (2012), I also use additional specifications of equations (1) and (2) in which I replace popularity and bargaining power with the residuals of two OLS estimates where these two measures are regressed on my performance variables (grades, goals and assists), plus age, age squared and season fixed effects.

We estimate equations (1) and (2) with OLS and panel estimators and considering random (RE) and fixed effects (FE) models. As known, fixed effect estimates allow us to take into account the effect of time-invariant players' unobservable characteristics (e.g., charisma, innate ability and ability to interact with other players) that could otherwise bias the estimates of the effect of my main variables of interest on players' salaries. Importantly, fixed effect estimates do not rely on the rather strong assumption of no correlation between individual time-invariant characteristics and earnings. However, I cannot rely entirely on fixed effects estimates in my context, as bargaining power is essentially time invariant because players change agents very rarely during their careers (as mentioned in Section 3). For this reason, I will employ fixed effect estimates to assess the impact of time-varying determinants (i.e., popularity and performance), while I will use both pooled OLS estimates and random effect estimates (which rely on the assumption of no correlation between time-invariant player characteristics and earnings) to assess the role of time-unvarying determinants (agent's power). However, all techniques lead us to draw similar conclusions with respect to the impact of time-varying determinants, thus providing robustness to my empirical strategy (see Section 5 for more details).

4.2 Unconditional Quantile Regression

To assess how the influence of the main determinants of super-earnings change along the earnings distribution – and especially at the top of the distribution, where "superstars" should lie –I apply models (1) and (2) on the pooled sample of players using the Unconditional Quantile Regression (UQR) approach, also called the Recentered Influence Function (RIF) method, proposed by Firpo et al. (2009).

The key advantage of the UQR approach over other distributional methods (i.e., the conditional quantile regression proposed by Koenker and Bassett 1978)[14] is that it allows us to analyse the relationship between covariates and the unconditional distribution of earnings. This possibility occurs because the UQR method provides a linear approximation of the unconditional quantiles of the dependent variable. The law of iterated expectations can be applied to the quantile being approximated and used to estimate the marginal effect of a covariate through a simple regression of a function of the outcome variable, the Recentered Influence Function, on the covariates.

In my setting, the RIF of earnings is estimated directly from the data by first computing the sample quantile $\boldsymbol{q}$ and then estimating the density of the distribution of income at that quantile using kernel density methods. Then, for a given observed quantile $\boldsymbol{q_\tau}$, a RIF is generated, which can take one of two values depending on whether the observation's value of the outcome variable is less than or equal to the observed quantile:

$$\boldsymbol{RIF\ (W; q_\tau) = q_\tau + \frac{\tau - 1[W \leq q_\tau]}{f_w(q_\tau)}\ (3)}$$

where $\boldsymbol{q_\tau}$ is the observed sample quantile of earnings, $\boldsymbol{\tau - 1[W \leq q_\tau]}$ is an indicator variable equal to one if the observation's value of earnings is less than or equal to the observed quantile and zero otherwise, while $\boldsymbol{f_w(q_\tau)}$is the estimated kernel density of earnings at the $\boldsymbol{\tau_{th}}$ quantile.

The RIF defined in equation (3) is then used as a dependent variable in an OLS regression on the covariates defined in equations (1) and (2). In practice, this amounts to estimate a rescaled linear

[14] Determinants of super-earnings in professional soccer are estimated through conditional quantile regressions by Lehmann and Schulze (2008), Franck and Nuesch (2012) and Deutsher and Buschemann (2016), while Deutscher et al. (2016) make use of unconditional quantile regressions.

probability model (Jones et al. 2015). Indeed, the unconditional quantile of earnings $\boldsymbol{q_\tau}$, may be obtained as follows:

$$\boldsymbol{q_\tau = E_x\left[E\left[\widehat{RIF}(W; q_\tau)|X\right]\right]} \ \textbf{(4)}$$

where $\boldsymbol{\widehat{RIF}(W; q_\tau)|X}$ is the estimate of RIF as defined in equation (3), conditional on covariates *X*. Thanks to this linear approximation, it is now possible to apply the law of iterated expectations. Thus, $\boldsymbol{q_\tau}$ can be written as

$$\boldsymbol{q_\tau = E[X]\widehat{\delta_\tau}} \ \textit{(5)}$$

where $\boldsymbol{\widehat{\delta_\tau}}$ is the coefficient of the unconditional quantile regression. This linearisation allows the estimation of the marginal effect of a change in distribution of covariates *X* on the unconditional quantile of earnings, measured by the parameter $\boldsymbol{\widehat{\delta_\tau}}$.

5. Results

5.1 OLS estimates

In Table 4, I report OLS estimates of equations (1) and (2), respectively, i.e., without and with team fixed effects.[15] The two equations are estimated including both rough measures of popularity and power (in the "baseline" model) and including nonperformance-related measures of popularity and power, as explained in Section 4. All coefficients are expressed in terms of one standard deviation (S.D.) of the variable. Therefore, with the dependent variable expressed in logs, estimated coefficients indicate the percentage change in annual net wages associated with a one-standard-deviation increase in the independent variable.

Estimates of equation (1) (columns 1 and 3, Table 4) show that all measures of performance, popularity and bargaining power exert a largely positive and highly statistically significant influence on earnings. In the baseline model (column 1, Table 4), for instance, a one-S.D. increase in goals, assists and mean grade during the previous season is associated with a wage increase of 11.4%, 3.4% and 6.3%, respectively. Likewise, a one-S.D. increase in proxies for popularity and bargaining power leads to a wage increase of 16% and 8.1%, respectively.

[15] Note that, for the sake of space, in this article, I show only estimated coefficients of my variables of interest (i.e., proxies of performance, popularity and power). Detailed estimates including all covariates are available upon request by the authors.

Interestingly, results are similar in magnitude and have the same level of statistical significance regardless of whether rough or non-performance-related measures of popularity and power are used as determinants of earnings.[16]

[Table 4 approximately here]

The size of the estimated coefficients decreases when team fixed effects are controlled for (columns 2 and 4, Table 4), but all coefficients remain statistically significant in models where nonperformance-related popularity and power measures are used. These results suggest that a pure compositional effect is at work – favouring players endowed by better talent, popularity or power to belong to a richer team – but all the determinants also have a direct effect on earnings (i.e., a "within team" effect). The largest drop in coefficients when team fixed effects are included refers to the proxy of bargaining power: the influence on annual net earnings of a one-S.D. increase in the player's portfolio owned by the agent drops from 8.1% to 2.1% when team fixed effects are added to the covariates. This is consistent with the role of the agent, which is essentially that of allocating the player to the team by guaranteeing the highest possible wage to the player.

5.2 Panel estimates

In table 5, I report random effects estimates according to all specifications discussed so far. The estimates basically confirm the results obtained through OLS, signalling that all three determinants of super-earnings also play a significant role when unobserved time-invariant individual characteristics (assumed to be independent with earnings, as in RE models) are taken into account (Columns 1 and 3 of Table 5). Moreover, the results are stable when "within team" estimates are carried out (Columns 2 and 4) and with both raw and nonperformance-related measures of popularity and bargaining power.

[Table 5 approximately here]

In Table 6, I report fixed effects estimates that do not include a time-invariant power variable, but do not rely on the assumption of no correlation between time-invariant individual characteristics and earnings, as discussed in Section 3. Fixed effect results differ with respect to the OLS and RE estimates in two ways. First, the size of all estimated coefficients decreases. Furthermore, among performance variables, both mean grade and goals are positively associated with earnings, but only goals remain

[16] The sample size of the regressions is lower than the total sample size because, using lagged covariates, I cannot include in the regressions those players who play for the first time in Italian *Serie A* in a certain season (for instance, foreign players have missing values for the first season they play in Italy).

statistically significant. Additionally, the proxy of popularity remains statistically significant and its size does not change when team fixed effects are added to the regression. For instance, according to the specification in Column 3, I find that a one-S.D. increase in the proxy of popularity brings about a remarkable 2.7% increase in earnings, while a one-S.D. increase in scored goals leads to a 5.0% increase in annual wages.

[Table 6 approximately here]

Overall, these results suggest that unobserved time-invariant player characteristics have a significant effect on earnings, and the size of the determinants of earnings is actually smaller when these characteristics are taken into account. However, both within-individual changes over time in goals – among other performance variables – and popularity highly contribute to increasing earnings when unobserved heterogeneity at both individual and team levels is taken into account. This suggests a robust causal link between these determinants and earnings.

5.3 Additional Checks

In this section, I perform a number of additional checks to test the robustness of my proxy of the agent's bargaining power. In fact, one peculiarity of the football market is that the agent negotiates the individual contract, so the single player might get benefits from being represented by a more powerful agent according to his client's portfolio. However, agents can have different business strategies in building their portfolio. To take into account this aspect, I exploit all the information available in my dataset by adding a number of additional and alternative measures of bargaining power in my baseline estimates. Results of these checks for both OLS and random effect specifications are reported in Tables A1 and A2, in the Appendix, respectively.

Firstly, in columns 1-2, I account for the size of agent's portfolio, adding the number of players represented by each agent as an additional control variable. The results related to my main variable (aggregated market value) are qualitatively unaffected, while the magnitude is only a bit higher compared to baseline estimates. The coefficient related to the number of players is very close to zero and barely statistically significant. Interestingly, in both OLS and random effect estimates, the effect of the aggregate market value is lower when team fixed effects are included as a control. This is in line with my baseline results and confirms the main role of the agent that is the one of allocating players to teams by guaranteeing the highest possible wage to the player.

Secondly, in columns 3-4, as an alternative bargaining power measure, I use the mean market value of the players in the agent's portfolio. This variable combines both information discussed above, i.e. the aggregate market value divided by the number of players represented by the same agent. Here again, my results are substantially unaffected and exhibit the same pattern discussed above when team fixed effects are also included.

Lastly, in columns 5-6, I account for the agent's portfolio diversification, adding a measure of the payers' value dispersion as an additional control. Unfortunately, for what concerns the market value, I only have aggregated data for each agent and detailed information on the portfolio are not available for the period to which my analysis refers to. However, I use the standard deviation of the wages of the Italian *Serie A* players represented by the same agent as an additional control. This measure is likely to capture agent's portfolio diversification, since the wage is a good proxy of player's market value (as shown by other papers, such as Forrest and Simmons 2002, Torgler and Schmidt 2007, Franck and Nüesch 2012). Also in this case, the inclusion of this additional control does not affect substantially my results. Only the aggregated market variable turns to be poorly statistical significant when team fixed effects are also included in the regression, in line with the results found above. The coefficient of this additional control is instead positive and significant meaning that portfolio diversification of the agent –other than aggregated portfolio value- is also positively correlated with player's earnings.

5.4 UQR estimates

Table 7 and Figure 3 show the results of UQR according to the specification of equation (1), while Table 8 and Figure 4 show UQR estimates of the specification including team fixed effects (as in equation (2)). Tables 7 and 8 show estimated coefficients at the 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles of the earnings distribution, respectively, while figures 3 and 4 show coefficients at every 5 percentiles of the distribution. For the sake of brevity, I show only the results of models where nonperformance-related measures of popularity and power are employed (alternative estimates are shown in Tables A3 and A4 in the Appendix and lead to similar results).

Table 7 and Figure 3 show that talent, popularity and bargaining power remain substantially positive and significant along the whole earnings distribution, and their size grows along the distribution. However, focusing on top percentiles, where superstar effects should more clearly manifest, the increasing pattern of popularity coefficients along the distribution must be emphasised. Popularity dominates all the other covariates in the top tail of the distribution and reaches its peak at the 95^{th} percentile, where a one-S.D. increase in popularity is associated with a 35.4% increase in annual earnings in the subsequent season. The effect of goals and grades on earnings is rather constant, and a one-S.D. increase in goals is associated with a 21.0% increase in annual earnings at the 95^{th} percentile of the earnings distribution. Conversely, bargaining power exerts a large and significant influence on earnings only up to the 75th percentile of the distribution, and it is not significantly associated with earnings after this threshold.

[Table 7 approximately here]

[Figure 3 approximately here]

Table 8 and Figure 4 show essentially the same pattern depicted in Table 7 and Figure 3, but with one important difference. Indeed, the inclusion of team fixed effects greatly reduces the influence of the agent's power, whose effect turns to be not statistically significant just after the median of the earnings distribution. A similar pattern, only observed at the mean of the earnings distribution, was also found in OLS estimates (shown in Section 5.1). Conversely, the influence of popularity is also large and significant when team fixed effects are included in the regression and its influence on earnings grows along their distribution: a one-S.D. increase in popularity is associated with a 4.1% increase in earnings at the median of the earnings distribution and with a 31.9% increase at the 95th percentile.

[Table 8 approximately here]

[Figure 4 approximately here]

Overall, UQR estimates generally support the results obtained from the OLS and panel regressions when these are evaluated at the mean of the distribution. However, UQR regression results reveal that the relative weight of these factors changes along the distribution of earnings, especially at the top tail of this distribution, where the effect of popularity dominates all the others.

6. Conclusions

The economic literature has addressed the phenomenon of the working super-rich by offering explanations based on individuals' talent, popularity or power. However, due to the difficulty of finding good proxies, no empirical studies have inquired about the joint influence of these factors on "super-earnings". Following suggestions by Kahn (2000) to exploit sports data in order to carry out innovative labour market research, I use a unique panel dataset on earnings and several characteristics of all Italian *Serie A* football players to investigate the joint effect of performance, popularity and bargaining power on earnings. An original feature of my dataset is the use of new and detailed information on players' performance (based on the goals, assists and grades given by leading Italian newspapers), popularity (based on the yearly searches on Google) and bargaining power (based on an agent's client portfolio), while the longitudinal nature of my dataset allows us to investigate the determinants of earnings while controlling for time-invariant players and club characteristics. Moreover, I use various estimation techniques to assess these relationships and, in particular, I employ Unconditional Quantile Regression to assess the role of the determinants along the entire distribution of players' earnings.

We find that proxies of talent, popularity and bargaining power are all significantly associated with higher earnings. However, I find that the relative weight of these factors on earnings greatly varies when considering a compositional vs a pure direct effect, and at different points along the earnings distribution. my main findings can be summarised in greater detail as follows.

First, according to my OLS estimates, I find that the leading determinants of earnings are represented by performance and popularity. A S.D. increase in goals, assists and mean grade during the previous season is associated with a wage increase of 11.4%, 3.4% and 6.3%, respectively. Similarly, a one-S.D. increase in annual searches of the player on Google is associated with a yearly earnings premium of approximately 16%. As a third determinant, I find that a one-S.D. increase in the agent's client portfolio, as a proxy of higher bargaining power, leads to an increase in earnings of approximately 8%, on average.

Second, I find that these results are partly due to a compositional effect – i.e., allocation of players endowed with greater talent, popularity and power to better teams – and partly by a direct influence on earnings. In particular, I find that an agent's market power is important, especially in allocating players to teams guaranteeing higher wages – a compositional effect – while its effect is very small when team fixed characteristics are taken into account.

Third, panel fixed effect estimates suggest that time-invariant individual characteristics have a significant effect on earnings, while within-individual changes over time in goals – among other performance variables – and popularity also highly contribute to increasing earnings. This suggests a robust causal link between these determinants and earnings.

Finally, I find through UQR that the relative weight of these determinants on earnings highly varies along the distribution of earnings, while analysis only at the mean seems to underestimate the differences in the wage structure between 'normal' players and 'superstars'. In particular, I find that popularity dominates all the other covariates at the top tail of the earnings distribution and reaches its peak at the 95^{th} percentile of this distribution. An increase in a player's popularity is magnified by an earning premium of approximately 35% at its maximum point, according to my estimates. Conversely, the role of bargaining power reaches its peak around the 75^{th} percentile, generating an earnings premium of approximately 14%, but it is not statistically associated with the earnings of players above this threshold.

The effect of performances is, instead, more constant and becomes relatively less important after the 75th percentile.

Overall, my findings suggest that the interpretations of extraordinary earnings based only on very talented workers who "win and take all" seems insufficient in order to wholly capture mechanisms behind top earners and that other mechanisms need to be taken into account. my findings, for instance, suggest that bargaining power plays a non-negligible role, and in the case of players, an agent's market power is important in order to negotiate better deals with team owners. The importance of this factor has also been found for other high earners, i.e., CEOs, and this encourages further research on the mechanism linking power and earnings and for other categories of the super-rich. my findings also suggest that popularity – above all – allows individuals to become super-rich, especially in a context – such as football – characterised by the large spread of pay-TV technology and the internet. This allows teams to be watched by a global audience and contributes to redistributing the largest share of revenues towards the most popular players. With a few possible differences across sectors, this can be a factor explaining the earnings of other high earners, such as actors, musicians, and of virtually all workers in sectors characterised by a large audience.

Future lines of research might benefit from the approach followed in this article jointly analysing the role of performance, popularity and power in the earnings of workers in sectors different from professional team sports. Moreover, a focus on the role of these determinants along the entire distribution of earnings seems to represent a promising strategy to better understand why superstars are paid so much.

References

Adler, M. (1985), Stardom and talent, *American Economic Review,* 75: 208-212.

Adler, M. (2006), Stardom and talent, in Edited by Ginsburgh, V. A., Throsby, D. (eds), *Handbook of the Economics of Art and Culture, Volume 1*, Elsevier.

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E. (2013), The top 1% in international and historical perspectives, *Journal of Economic Perspectives*, 27: 3–20.

Atkinson, A.B., Piketty, T., Saez, E. (2011), Top incomes in the long run of history, *Journal of Economic Literature*, 49: 3–71.

Bebchuck, L. A., Fried, J. M. (2003), Executive Compensation as an Agency Problem, *Journal of Economic Perspectives*, 17: 71-92.

Bivens, J., Mishel, L. (2013), The pay of corporate executives and financial professionals as evidence of rents in top 1 percent, *Journal of Economic Perspectives*, 27: 57–78.

Blair, R.D. (2012), *Sports Economics*, Cambridge University Press.

Boeri, T., Severgnini, B. (2012), The Decline of Professional Football in Italy, *IZA Discussion Paper*, n. 7018.

Brandes, L., Franck, E., Nüesch, S. (2008), Local Heroes and Superstars An Empirical Analysis of Star Attraction in German Soccer, *Journal of Sports Economics*, 9: 266-286.

Bryson, A., Frick, B., Simmons, R. (2012), The Returns to Scarce Talent Footedness and Player Remuneration in European Soccer, *Journal of Sports Economics*, 14: 606-628.

Buraimo, B., Frick B., Hickfang, M., Simmons R. (2015), The economics of long-term contracts in the footballers' labour market, *Scottish Journal of Political Economy*, 62: 8-24.

Carmichael, F., Rossi, G., Simmons, R. (2012). Contract Duration and Player Performance in Italian Football, *4th European Conference in Sport Economics / XIV IASE*, Birkbeck University of London.

Deutscher, C., Buschemann, A. (2016), Does Performance Consistency Pay Off Financially for Players? Evidence From the Bundesliga, *Journal of Sports Economics*, 17: 27-43.

Deutscher, C., Gürtler, O., Prinz, J., Weimar, D. (2016), The Payoff To Consistency In Performance, *Economic Inquiry*, doi: 10.1111/ecin.12415.

Dickson, A., Jennings, C., & Koop, G. (2016). Domestic violence and football in Glasgow: are reference points relevant?, *Oxford Bulletin of Economics and Statistics*, *78:* 1-21.

Firpo, S., Fortin, N., Lemieux, T., (2009), Unconditional Quantile Regressions, *Econometrica*, 77: 953-973.

Forrest, D., & Simmons, R. (2002). Team salaries and playing success in sports: a comparative perspective. In *Sportökonomie* (pp. 221-238). Gabler Verlag, Wiesbaden.

Franck, E., Nüesch, S. (2006), Talent, Past Consumption and/or Popularity - Are German Soccer Celebrities Rosen or Adler Stars?, *Working Papers 0043, University of Zurich, Institute for Strategy and Business Economics* (ISU).

Franck, E., Nüesch, S. (2012), Talent and/or Popularity - What Does it Take to Be a Superstar?, *Economic Inquiry*, 50: 202-216.

Frank, R.H., Cook, P.J. (1995), *The winner-take-all society*, Free Press, New York.

Franzini, M., Granaglia, E., Raitano, M. (2016), *Extreme Inequalities in Contemporary Capitalism: ShouldIBe Concerned About the Rich?*, Springer.

Frick, B. (2007), The football players' labor market: empirical evidence from the major european leagues, *Scottish Journal of Political Economy*, 54: 422-446.

Gallo, E., Grund, T., & James Reade, J. (2013). Punishing the foreigner: implicit discrimination in the Premier League based on oppositional identity, *Oxford Bulletin of Economics and Statistics*, *75*: 136-156.

Jones, A., Lomas, J., Rice, N. (2015), Healthcare Cost Regressions: Going Beyond the Mean to Estimate the Full Distribution, *Health Economics*, 24: 1192-1212.

Kahn, L.M. (2000), The Sports Business as a Labor Market Laboratory, *Journal of Economic Perspectives*, 14: 75-94.

Koenker, R., Basset, G. (1978), Regression quantiles, *Econometrica*, 46: 33–50.

Krueger, A., B. (2005), The economics of Real Superstars: The Market for Rock Concerts in the Material World, *Journal of Labor Economics*, 23: 1-30.

Lehmann, E. E., Schulze, G. G. (2008), What does it take to be a star? The role of performance and the media for German soccer players, *Applied Economics Quarterly*, 54: 59-70.

Lucifora, C., Simmons, R. (2003), Superstar Effects in Sports: Evidence from Italian Soccer, *Journal of Sports Economics,* 15: 35-55.

Mason, D. (2012), Player agents, in Andreff, W., Szymanski, S. (eds.), *Handbook on the Economics of Sport*, Edward Elgar Publishing.

Mincer, J. (1974), *Schooling, Experience and Earnings*, New York, Columbia University Press.

Mullin, C. J., Dunn, L. F. (2002), Using baseball card prices to measure star quality and monopsony, *Economic Inquiry*, 40: 620-632.

Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees, *The Quarterly Journal of Economics*, *125*: 1859-1887.

Reilly, B., Witt, R., (2007). The Determinants of Base Pay and the Role of Race in Major League Soccer: Evidence from the 2007 League Season, *School of Economics Discussion Papers 1907*, School of Economics, University of Surrey.

Rosen, S. (1981), Economics of superstars, *American Economic Review,* 71: 167-183.

Rosen, S., Sanderson, A. (2001), Labour Markets in Professional Sports, *The Economic Journal,* 111: F47-F68.

Send. J. (2016), Football and Money: Income Inequality in the German Bundesliga, https://the10thmanblog.wordpress.com/

Torgler, B., & Schmidt, S. L. (2007). What shapes player performance in soccer? Empirical findings from a panel analysis, *Applied Economics*, *39*: 2355-2369.

Treme, J., Allen, S. K. (2009), Widely Received: Payoffs to Player Attributes in the NFL, *Economics Bulletin,* 29: 1631-1643.

Zimbalist, A. (1992), *Baseball and Billions*, New York: Basic Books.

Tables and Figures

Table 1. Sample characteristics

Variable	Description	Mean (St. dev.)
Dependent variable		
Wage	Net earnings (pre-season values) in thousands/€	875.1 (911.7)
Log wage	Log of net earnings (pre-season values)	6.38 (0.87)
Individual controls		
Age	Age (years)	26.6 (4.2)
Age square	Age squared	725.8 (226.7)
Position	Dummies for defenders (40.2%), midfielder (39.9%) and forward (19.9%)	
Captain	Dummy for the team's captain	0.034 (0.181)
Minutes played	Minutes played during the season	1352.8 (1068.3)
Total international caps	Number of caps with the national team up to 2014-2015	15.75 (25.10)
Total Under-21 caps	Number of caps with the U21 national team up to 2014-2015	5.85 (8.68)
International caps	Number of caps with the national team during the season	2.01 (4.65)
Under-21 caps	Number of caps with the U21 national team during the season	0.37 (1.59)
Player's performance		
Grade	Mean grade by newspapers during the season	5.77 (0.41)
Goal	Goal scored during the season	1.93 (3.50)
Assist	Assist served during the season	1.28 (2.04)
Measure of popularity		
Popularity	Google researches results (million)	4.21 (9.37)
Measures of power		
Agent Market value	Market value of players represented by the same agent (in millions/€)	101.6 (214.4)
Number of players	Number of players represented by the same agent	47.4 (104.7)
Mean market value	Mean market value of players represented by the same agent (in millions/€)	1.83 (2.23)
Wage standard deviation	Standard deviation of players' wages in agent's portfolio (in thousands/€)	348.8 (459.7)

Table 2. Distributions of annual net wages: percentiles, percentile ratios and Gini coefficient

	2013-2014	All seasons
Mean	787.2	875.1
Standard Deviation	877.2	911.7
Minimum	30	30
P10	200	200
P25	300	300
P50	500	550
P75	900	1000
P90	1700	2000
P95	2500	3000
Maximum	6500	6500
P95/P90	1.5	1.5
P95/P75	2.8	3.0
P95/P50	5.0	5.5
P95/P25	8.3	10.0
P95/P10	12.5	15.0
Gini coefficient	0.490	0.474

[a] Expressed in thousands of Euros

Table 3. Distribution of team's characteristics in 2013-2014[a]

	Net Sales	Earnings before taxes	Revenues from TV rights	Revenues from games tickets
Mean	92,815	-8,263	54,808	10,494
S.D.	72,778	24,918	37,895	10,629
Minimum	34,348	-79,882	25,164	1,516
p10	42,318	-36,740	29,552	1,956
p25	44,724	-14,040	29,870	3,903
p50	56,215	-2,796	33,937	4,187
p75	116,446	1,441	66,014	15,134
p90	246,679	14,261	119,547	28,698
Maximum	272,404	44,124	163,478	38,051

[a] Only teams who also participated in the *Serie A* during the 2012-2013 season are considered

Table 4. Association between annual net (log) wages, performance, popularity and power. OLS estimates

	Baseline		Nonperformance-related measures of popularity and power[a]	
	No team fixed effects[b]	Team fixed effects[c]	No team fixed effects[b]	Team fixed effects[c]
grade	0.0630***	0.0308**	0.0792***	0.0392***
	0.0207	*0.0148*	*0.0206*	*0.0148*
goal	0.1144***	0.0637***	0.1231***	0.0690***
	0.0224	*0.0161*	*0.0224*	*0.0161*
assist	0.0339*	0.0184	0.0519***	0.0286**
	0.0185	*0.0131*	*0.0185*	*0.0131*
popularity	0.1599***	0.1000***		
	0.0198	*0.0143*		
power	0.0812***	0.0214*		
	0.0170	*0.0124*		
popularity[a]			0.1338***	0.0837***
			0.0166	*0.0120*
power[a]			0.0770***	0.0203*
			0.0161	*0.0117*
Obs.	1198	1198	1198	1198

[a]Ireplace among regressors popularity and power measures with the residuals of OLS estimates on popularity and power, respectively, ran including among regressors goal, assist, grade, age, age squared and season dummies. [b] The following control variables are included (all referred to the previous season): age and age squared; dummies on citizenship (Italian, EU, extra EU); dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season and until 2014-2015 season; number of national under-21 team caps during the season and until 2014-2015 season; season fixed effects. [c] Team fixed effects are added to the control variables of the baseline model. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Association between annual net (log) wages, performance, popularity and power.
Random effects estimates

	Baseline		Nonperformance-related measures of popularity and power[a]	
	No team fixed effects[b]	Team fixed effects[c]	No team fixed effects[b]	Team fixed effects[c]
grade	0.0307**	0.0237*	0.0410***	0.0309**
	0.0145	*0.0128*	*0.0146*	*0.0128*
goal	0.0772***	0.0594***	0.0811***	0.0632***
	0.0181	*0.0156*	*0.0181*	*0.0156*
assist	0.0116	0.0105	0.0214	0.0184
	0.0141	*0.0122*	*0.0141*	*0.0122*
popularity	0.0684***	0.0692***		
	0.0153	*0.0132*		
power	0.0947***	0.0393**		
	0.0257	*0.0176*		
popularity[a]			0.0573***	0.0579***
			0.0128	*0.0111*
power[a]			0.0898***	0.0372**
			0.0244	*0.0167*
Obs.	1198	1198	1198	1198

[a]Ireplace among regressors popularity and power measures with the residuals of OLS estimates on popularity and power, respectively, ran including among regressors goal, assist, grade, age, age squared and season dummies. [b] The following control variables are included (all referred to the previous season): age and age squared; dummies on citizenship (Italian, EU, extra EU); dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season and until 2014-2015 season; number of national under-21 team caps during the season and until 2014-2015 season; season fixed effects. [c] Team fixed effects are added to the control variables of the baseline model. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Association between annual net (log) wages, performance and popularity.
Fixed effects estimates

	Baseline		Nonperformance-related measures of popularity[a]	
	No team fixed effects[b]	Team fixed effects[c]	No team fixed effects[b]	Team fixed effects[c]
grade	0.0192	0.0127	0.0216	0.0151
	0.0149	*0.0134*	*0.0149*	*0.0134*
goal	0.0481**	0.0351**	0.0498***	0.0369**
	0.0190	*0.0170*	*0.0191*	*0.0171*
assist	-0.0020	-0.0074	0.0011	-0.0043
	0.0145	*0.0129*	*0.0145*	*0.0129*
popularity	0.0323**	0.0334**		
	0.0160	*0.0142*		
popularity[a]			0.0271**	0.0280**
			0.0134	*0.0119*
Obs.	1198	1198	1198	1198

[a]Ireplace among regressors popularity measure with the residuals of OLS estimates on popularity ran including among regressors goal, assist, grade, age, age squared and season dummies. [b]The following control variables are included (all referred to the previous season): age and age squared; dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season; number of national under-21 team caps during

the season; season fixed effects. [c] Team fixed effects are added to the control variables of the baseline model. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Association between annual net (log) wages, performance, popularity and power. UQR[ab]

	Q10	Q25	Q50	Q75	Q90	Q95
grade	0.0352	0.0631**	0.1073***	0.1220***	0.1175***	0.0979*
	0.0297	*0.0277*	*0.0302*	*0.0451*	*0.0445*	*0.0511*
goal	-0.0021	0.0120	0.1466***	0.3346***	0.2170**	0.2100**
	0.0178	*0.0202*	*0.0323*	*0.0555*	*0.0865*	*0.0994*
assist	-0.0203	0.0249	0.0684**	0.1245***	0.1086**	0.0975
	0.0151	*0.0178*	*0.0286*	*0.0477*	*0.0530*	*0.0711*
popularity[a]	0.0255**	0.0374**	0.0923***	0.2189***	0.2788***	0.3540***
	0.0107	*0.0169*	*0.0252*	*0.0489*	*0.0546*	*0.0875*
power[a]	0.0312**	0.0507***	0.1283***	0.1380***	-0.0539	0.0024
	0.0154	*0.0160*	*0.0209*	*0.0470*	*0.0498*	*0.0618*
Team Fixed Effects	No	No	No	No	No	No
Obs.	1198	1198	1198	1198	1198	1198

[a]Ireplace among regressors popularity and power measures with the residuals of OLS estimates on popularity and power, respectively, ran including among regressors goal, assist, grade, age, age squared and season dummies. [b] The following control variables are included (all referred to the previous season): age and age squared; dummies on citizenship (Italian, EU, extra EU); dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season and until 2014-2015 season; number of national under-21 team caps during the season and until 2014-2015 season; season fixed effects. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Association between annual net (log) wages, performance, popularity and power. UQR[ab]

	Q10	Q25	Q50	Q75	Q90	Q95
grade	0.0215	0.0410*	0.0598**	0.0282	0.0625	0.0605
	0.0276	*0.0240*	*0.0269*	*0.0410*	*0.0432*	*0.0520*
goal	-0.0123	-0.0199	0.0807***	0.2185***	0.1567**	0.1651*
	0.0182	*0.0199*	*0.0283*	*0.0471*	*0.0742*	*0.0917*
assist	-0.0334**	0.0045	0.0336	0.0803**	0.0955*	0.1010
	0.0154	*0.0168*	*0.0249*	*0.0406*	*0.0509*	*0.0708*
popularity[a]	0.0137	0.0093	0.0409**	0.1054***	0.2070***	0.3189***
	0.0106	*0.0131*	*0.0196*	*0.0344*	*0.0485*	*0.0875*
power[a]	0.0131	0.0155	0.0463***	0.0259	-0.0957**	-0.0436
	0.0129	*0.0135*	*0.0176*	*0.0389*	*0.0483*	*0.0600*
Team Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1198	1198	1198	1198	1198	1198

[a]Ireplace among regressors popularity and power measures with the residuals of OLS estimates on popularity and power, respectively, ran including among regressors goal, assist, grade, age, age squared and season dummies. [b] The following control variables are included (all referred to the previous season): age and age squared; dummies on citizenship (Italian, EU, extra EU); dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season and until 2014-2015 season; number of national under-21 team caps during the season and until 2014-2015 season; season fixed effects; team fixed effects. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1. Kernel density estimate of annual net wages in 2013-2014

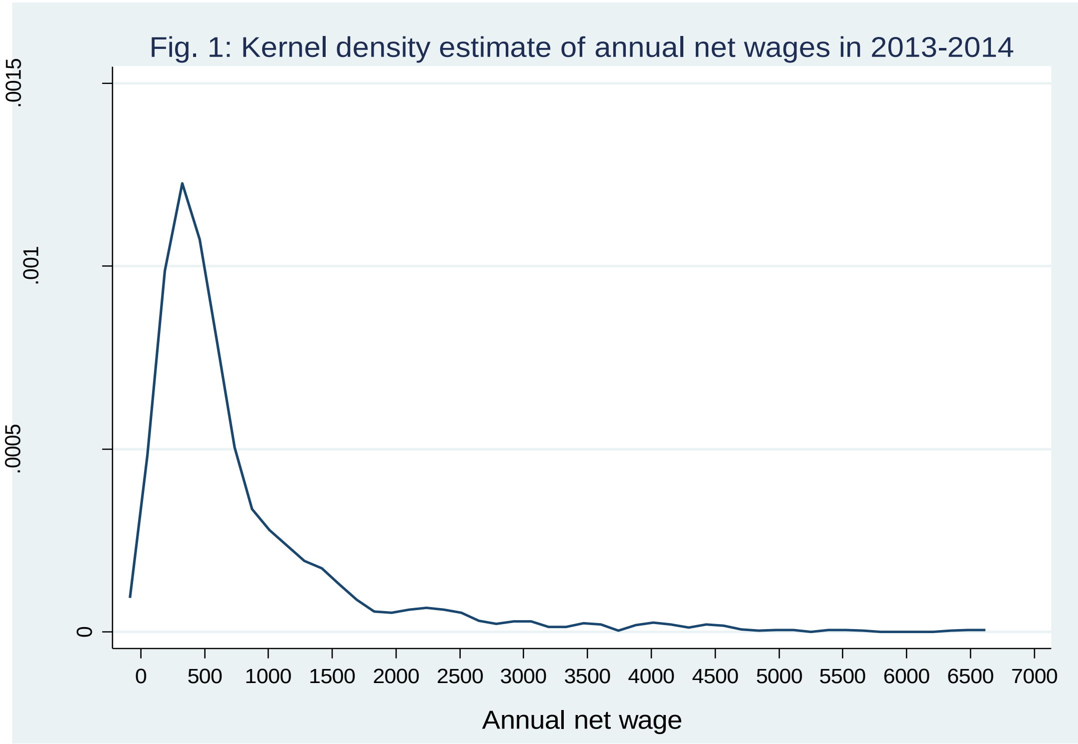
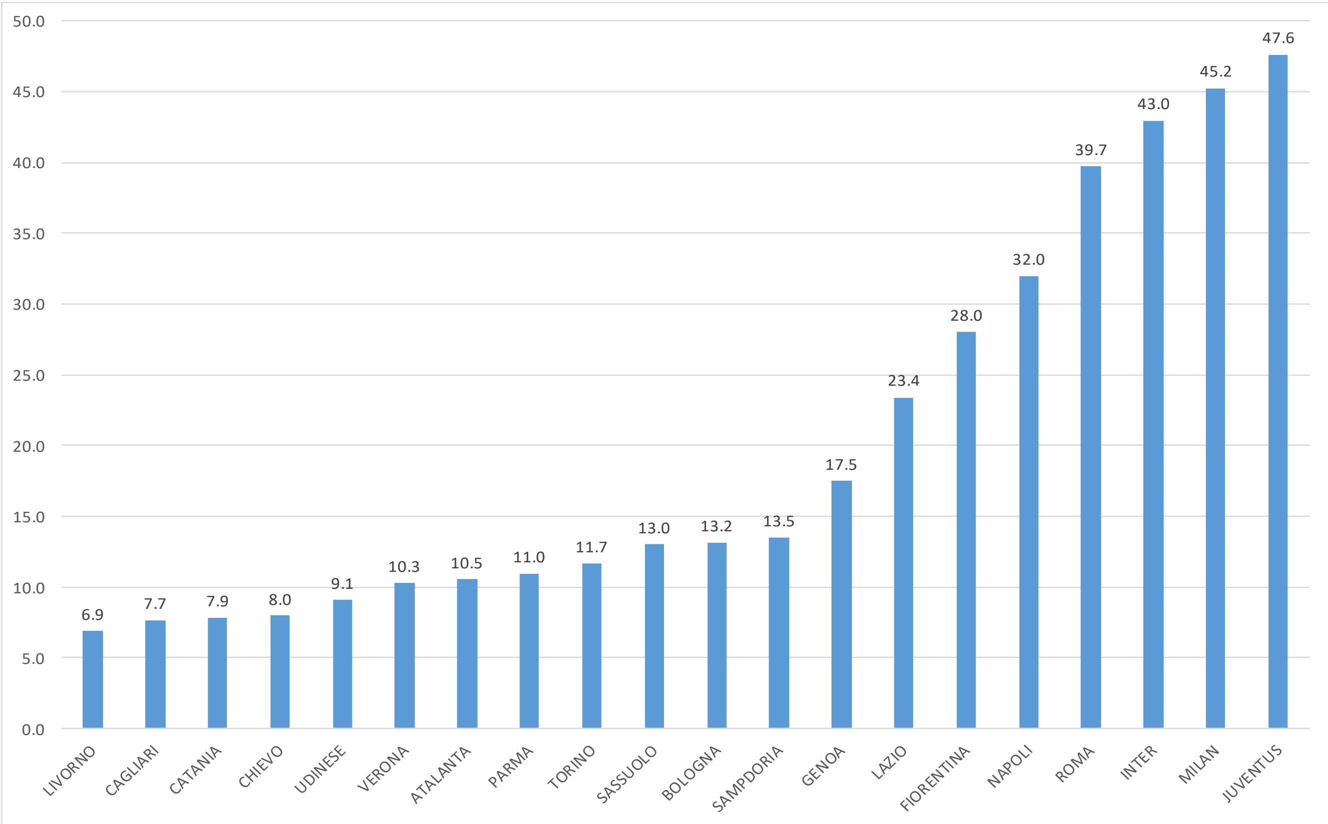
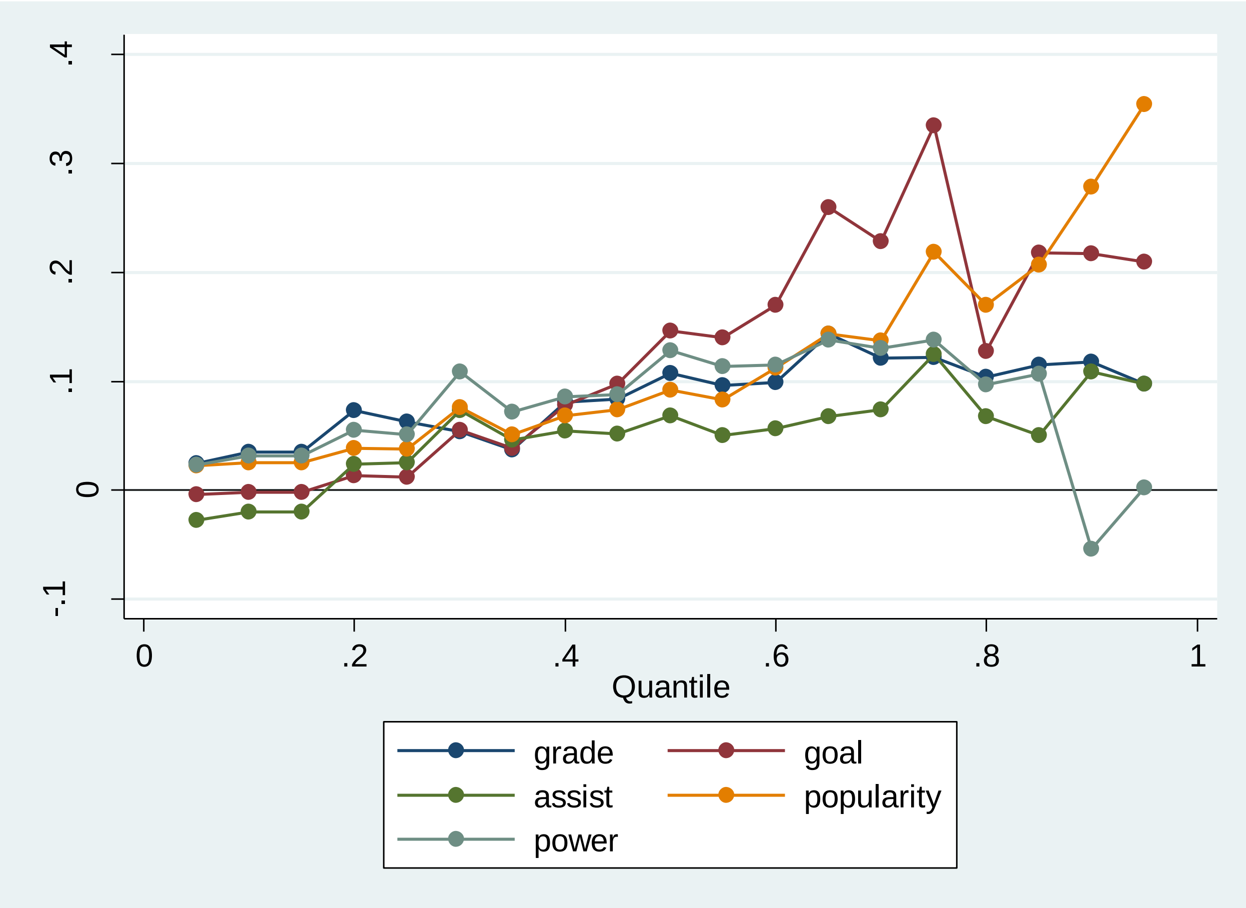


Figure 2. Total net wage bill paid to players by teams participating to the 2013-2014 season[a]



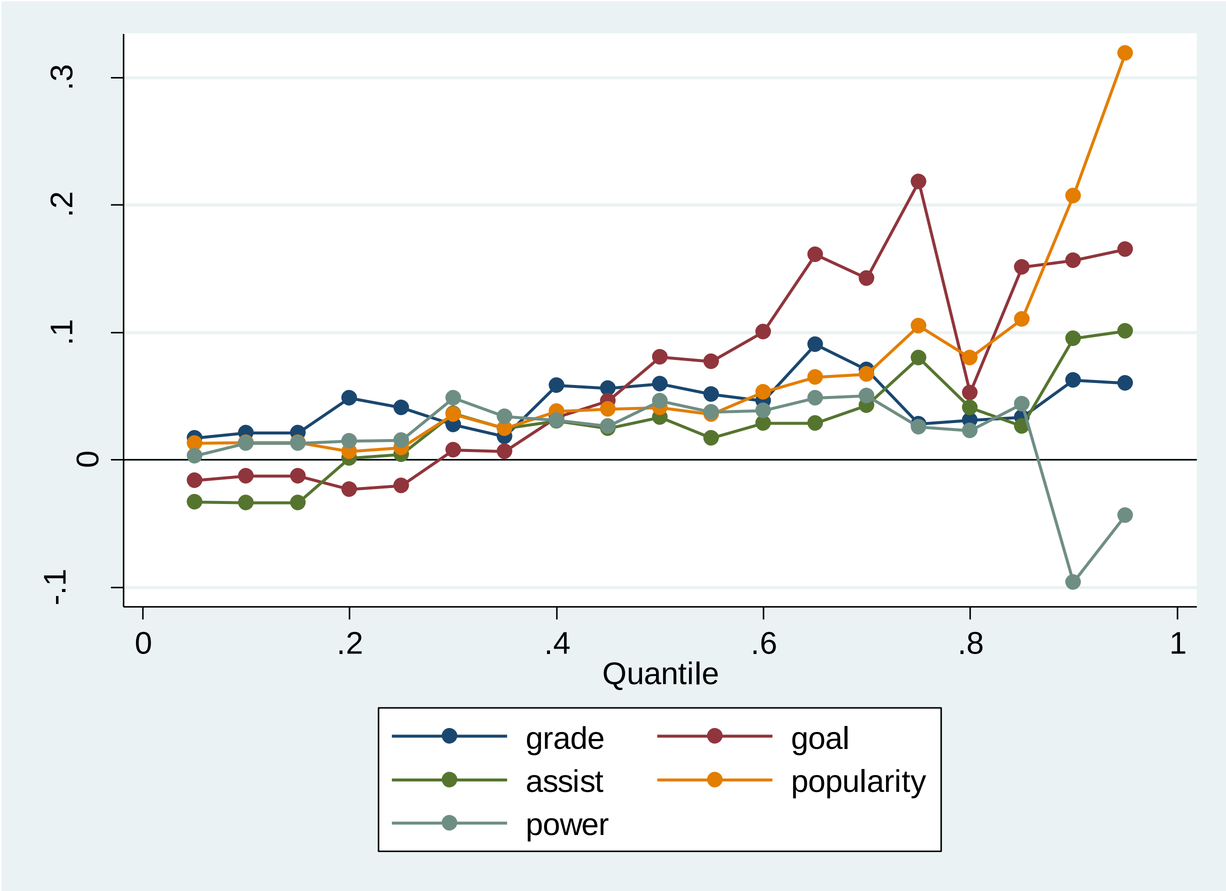
[a] Goalkeepers' earnings are included in the computation

Figure 3. Estimated coefficients of the association between annual net (log) wages, performance, popularity and power along the earnings distribution[a]. UQR- No team fixed effects model



[a]Nonperformance-related measures of popularity and power employed.

Figure 4. Estimated coefficients of the association between annual net (log) wages, performance, popularity and power along the earnings distribution[a]. UQR- Team fixed effects model



[a]Nonperformance-related measures of popularity and power employed.

Appendix

Table A1. Association between annual net (log) wages, performance, popularity and alternative measures of power.
OLS estimates

	No team fixed effects	Team fixed effects[b]	No team fixed effects	Team fixed effects[b]	No team fixed effects	Team fixed effects[b]
grade	0.0647***	0.0312**	0.0588***	0.0299**	0.0639***	0.0318**
	0.0206	*0.0148*	*0.0205*	*0.0148*	*0.0205*	*0.0148*
goal	0.1133***	0.0635***	0.1143***	0.0640***	0.1133***	0.0630***
	0.0224	*0.0161*	*0.0222*	*0.0160*	*0.0222*	*0.0160*
assist	0.0346*	0.0186	0.0319*	0.018	0.0355*	0.0192
	0.0184	*0.0131*	*0.0183*	*0.0130*	*0.0183*	*0.0130*
popularity	0.1545***	0.0990***	0.1424***	0.0950***	0.1504***	0.0967***
	0.0198	*0.0144*	*0.0197*	*0.0144*	*0.0197*	*0.0143*
Measures of Agent's bargaining power						
market value	0.1443***	0.0352*			0.0619***	0.0132
	0.0279	*0.0203*			*0.0173*	*0.0126*
number of players	-0.0008***	-0.0002				
	0.0003	*0.0002*				
mean market value			0.0502***	0.0152***		
			0.0072	*0.0054*		
wage standard dev.					0.0002***	0.0001***
					0.0000	*0.0000*
Obs.	1198	1198	1198	1198	1198	1198

[a] The following control variables are included (all referred to the previous season): age and age squared; dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season; number of national under-21 team caps during the season; season fixed effects. [b] Team fixed effects are added to the control variables of the baseline model. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table A2. Association between annual net (log) wages, performance, popularity and alternative measures of power.
Random effects estimates

	No team fixed effects	Team fixed effects[b]	No team fixed effects	Team fixed effects[b]	No team fixed effects	Team fixed effects[b]
grade	0.0308**	0.0238*	0.0290**	0.0226*	0.0304**	0.0239*
	0.0145	*0.0128*	*0.0144*	*0.0127*	*0.0145*	*0.0128*
goal	0.0767***	0.0593***	0.0760***	0.0590***	0.0772***	0.0593***
	0.0181	*0.0156*	*0.0180*	*0.0155*	*0.0180*	*0.0156*
assist	0.0112	0.0103	0.0106	0.0101	0.0120	0.0107
	0.0141	*0.0122*	*0.0140*	*0.0122*	*0.0141*	*0.0122*
popularity	0.0661***	0.0678***	0.0607***	0.0636***	0.0661***	0.0672***
	0.0153	*0.0132*	*0.0152*	*0.0132*	*0.0153*	*0.0132*
Measures of Agent's bargaining power						
market value	0.1778***	0.0734**			0.0690***	0.0260
	0.0413	*0.0287*			*0.0267*	*0.0184*
number of players	-0.0011**	-0.0004				
	0.0004	*0.0003*				
mean market value			0.0718***	0.0326***		
			0.0113	*0.0080*		
wage standard dev.					0.0002***	0.0001**
					0.0001	*0.0000*
Obs.	1198	1198	1198	1198	1198	1198

[a] The following control variables are included (all referred to the previous season): age and age squared; dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season; number of national under-21 team caps during the season; season fixed effects. [b] Team fixed effects are added to the control variables of the baseline model. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Association between annual net (log) wages, performance, popularity and power. RIF regressions.[a]

	Q10	Q25	Q50	Q75	Q90	Q95
grade	0.0312	0.0568**	0.0916***	0.0949**	0.0968**	0.0673
	0.0298	*0.0278*	*0.0304*	*0.0452*	*0.0443*	*0.0505*
goal	-0.0039	0.0094	0.1403***	0.3203***	0.1997**	0.1878*
	0.0178	*0.0201*	*0.0322*	*0.0554*	*0.0865*	*0.0995*
assist	-0.0243	0.0188	0.0534*	0.0947**	0.0789	0.0573
	0.0152	*0.0178*	*0.0285*	*0.0479*	*0.0536*	*0.0703*
popularity	0.0304**	0.0447**	0.1102***	0.2615***	0.3331***	0.4229***
	0.0128	*0.0202*	*0.0301*	*0.0585*	*0.0652*	*0.1045*
power	0.0329**	0.0535***	0.1353***	0.1456***	-0.0568	0.0025
	0.0163	*0.0169*	*0.0220*	*0.0496*	*0.0525*	*0.0651*
Team F.E.	No	No	No	No	No	No
Obs.	1198	1198	1198	1198	1198	1198

[a] The following control variables are included (all referred to the previous season): age and age squared; dummies on citizenship (Italian, EU, extra EU); dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season and until 2014-2015 season; number of national under-21 team caps during the season and until 2014-2015 season; season fixed effects. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. Association between annual net (log) wages, performance, popularity and power. RIF regressions.[a]

	Q10	Q25	Q50	Q75	Q90	Q95
grade	0.0195	0.0393	0.0535**	0.0176	0.0504	0.0357
	0.0276	*0.0241*	*0.0269*	*0.0409*	*0.0430*	*0.0516*
goal	-0.0132	-0.0205	0.0780***	0.2118***	0.1442*	0.1453
	0.0182	*0.0198*	*0.0282*	*0.0470*	*0.0741*	*0.0919*
assist	-0.0354**	0.0029	0.0273	0.0674*	0.0755	0.0664
	0.0153	*0.0168*	*0.0250*	*0.0406*	*0.0513*	*0.0698*
popularity	0.0164	0.0111	0.0489**	0.1259***	0.2473***	0.3810***
	0.0127	*0.0157*	*0.0234*	*0.0411*	*0.0579*	*0.1046*
power	0.0138	0.0164	0.0488***	0.0273	-0.1010**	-0.0460
	0.0136	*0.0142*	*0.0185*	*0.0410*	*0.0509*	*0.0633*
Team F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1198	1198	1198	1198	1198	1198

[a] The following control variables are included (all referred to the previous season): age and age squared; dummies on citizenship (Italian, EU, extra EU); dummies for the position on the pitch; dummy for team captain; number of played minutes; number of national team caps during the season and until 2014-2015 season; number of national under-21 team caps during the season and until 2014-2015 season; season fixed effects; team fixed effects. Standard Errors in italics. *** p<0.01, ** p<0.05, * p<0.1.

CHAPTER II

Health shocks and labour market outcomes: evidence from professional football

Abstract

This study uses traumatic injuries as a source of exogenous variation in professional football players' health combined with an IV strategy to provide estimates of the causal impact of a health shock on labour market outcomes. Using a unique longitudinal dataset on the universe of professional players in the Italian Serie A, I find that 30-days injury increases the probability of renegotiating the contract by about 10% and reduce net wages by around 12%. Compared to other results, my own demonstrate that health shocks can have a large impact on the earnings of super-rich especially among those largely relying on physical skills. This effect is largely driven by employer's concern about depreciation in the worker's human capital rather than by the shock-induced reduction in his productivity.

An extract from this chapter is published as: Carrieri, V., Jones, A. M., & Principe, F. (2018). Health shocks and labour market outcomes: evidence from professional football (No. 18/01). *HEDG*, c/o Department of Economics, University of York.

1. Introduction

Health shocks are a major concern for the labour market. In fact, health shocks affect peoples' lives across a number of different dimensions including productivity, labour market participation, labour income and healthcare costs, resulting in reduced household disposable income. The direct effect of health shocks on income and education is also the basis of the *health selection* hypothesis that is one of the leading explanations - along with the direct effect of income on health - of the socio-economic gradient in health observed in most countries (Deaton, 2003). The labour market consequences of health shocks have been the object of a large literature intersecting both health and labour economics with evidence from both developed and developing countries (see Section 2 for a literature review). However, due to the scarcity of relevant data, most of the contributions have focused on elderly workers and have analysed labour market participation as their main outcome.

Furthermore, to the best of my knowledge, there are no studies that analyse the effects of the health shocks on the top tail of the labour income distribution, the so-called "working super-rich". This segment of the distribution has been the subject of keen economic interest for a number of reasons. Firstly, since the 1970s, top income shares started to rise substantially in all developed countries, in particular, the English-speaking ones (Atkinson *et al.*, 2011). Secondly, the share of top incomes is considered a reliable indicator of long run inequality across the entire distribution. In fact, the income concentration at the top of the distribution is highly correlated with relative poverty and other inequality measures (Leigh, 2007). The analysis of income concentration at the top is then useful to understand also the evolution of the bottom of the income distribution. Lastly, following the recent contribution of Alvaredo *et al.* (2013), a substantial change in the composition of income shares at the top of the distribution has been found. In fact, in all the countries for which data are available, there has been a shift from capital income to labour income as the main component of the top 1% income shares. For instance, in Italy, earnings accounted for the 46.4% of the total in 1980 while they accounted for 70.9% of the total in 2008 and, similarly, among the 0.1% richest segment of the population, the share of earnings rose from 29.5% to 66.2% in the same period (Franzini et al. 2016). Thus, the labour market seems to be fertile ground for the escalation of contemporary society's extreme inequalities and, understanding how the "working super-rich" respond to health shocks can be a key element of these dynamics.

This chapter aims to fill this gap by providing evidence of the relationship between health shocks and labour market outcomes of individuals in the top tail of earnings distribution. In particular, the aim of this research is to analyse traumatic injuries as exogenous variation in professional football[17]

[17] Throughout this paper, I use the word "football" to indicate European football. In the US, this is known as "soccer".

players' health to provide estimates of the causal impact of a health shock on two main labour market outcomes: the annual net wages and the probability of renegotiating the contract between the employer (the club) and the employee (the player). To this end, I have created a unique longitudinal dataset recording data about wages, performances, popularity and injuries of the universe of football players of the Italian Serie A followed from 2009 to 2014[18]. This dataset was built by merging information from several reliable sources of data.

The choice of focusing on data from professional football is based on a number of considerations. Firstly, the football market is based on short-term contracts (generally 3-4 years), generating high volumes on renegotiation and players being traded in every market window; this allows for observations based on a large variation in wages across seasons. Secondly, the football players constitute a large share of the top earners. In fact, in 2003, among the top 500 earners, representing the top 0.01% of the distribution, about 20% were football players or managers in Italy (Franzini *et al.* 2016). Lastly, as argued by Kahn (2000): "professional sports offers a unique opportunity for research. There is no research setting other than sports where I know the name, face, and life history of every production worker and supervisor in the industry". In fact, professional sports data allow a perfect match between employer (the club) and employee (the player) and to have measures to proxy individual productivity and estimate team production functions. In recent years, the use of sports data to address health-related issues has been rising. Recent contributions include Stoecker *et al.* (2016), who examine the impact of influenza transmission on mortality by looking at local sports team success through the participation in Super Bowl of the National Football League (NFL)[19], and Hanson *et al.* (2017), who exploit a difference-in-differences framework in order to examine the effects of the "Crown of the Helmet Rule" on players' injuries in the NFL[20].

This study makes a number of contributions to the literature. Firstly, as previously discussed, it focuses on the consequences of health shocks for those on top incomes, for whom there is scant evidence so far. Secondly, it provides causal evidence of the relationship between health shocks and labour market outcomes. This is a rather difficult task given the possible endogeneity of health shocks in relation to labour supply, which may result from both simultaneity and unobserved heterogeneity (Lindeboom and Kerkhofs 2009; Cai, 2010). In this study, I exploit the exogenous (traumatic) nature of the health shock in which player may incur, the longitudinal nature of my dataset which allows us

[18] Italian Serie A is one of the five most followed football leagues in the world, alongside the Premier League (England), Bundesliga (Germany), La Liga (Spain) and Ligue 1 (France). Yearly, it has a turnover of about 1.9 billion Euros. Among the "big five" leagues, Italian Serie A is the one with the highest incidence of the wage bill on the clubs' costs, absorbing about 70% of the clubs' total earnings (Deloitte, 2017).

[19] American football league.

[20] Implemented in 2013 by NFL, it aims at reducing incidence of concussions and head injuries by penalizing a player who intentionally initiates contact with another player using the top of his helmet.

to sweep out individual time invariant characteristics and an instrumental variable strategy (using the average number of yellow cards received by the team as an instrument) to reduce any further endogeneity concerns. Lastly, my data allows for a deeper analysis of the main mechanisms of the health shock, disentangling the effect mediated through the player's performance and the one generated by human capital depreciation, inducing the club to offer a lower wage for precautionary reasons.

We find that health shocks affect significantly the labour market outcomes of super-rich but in a very particular way if compared to "ordinary" workers. First, in terms of magnitude, I find that a 30-day injury results in a reduction of the following year's wage by about 12%. This effect is significantly higher than the one found for other workers. Interestingly, I also find that this reduction can be explained more in terms of precautionary measures being taken by the club rather than through any direct effect that the injury might have had on productivity. Consistently, the injury is found to be positively associated with the probability of renegotiating the contract at the end of the season with the same club or by moving to another club. This is particularly true for more severe injuries. Finally, I find that the main factors addressed by the literature about the determinants of super-earnings, such as performance and popularity, have a significant impact on the players' wages. These results are robust to a number of model specifications and to alternative measures of health shocks and may be generalizable to other categories of super-rich whose earnings reflect competitive performance and popularity with the public.

The remainder of the chapter is structured as follows. Section 2 provides some insight about the literature on health shocks' consequences on employees' outcomes and the theoretical and empirical evidence about working super-rich's wages determination. Section 3 describes the data gathering process, explaining the main variables and providing some descriptive statistics. Section 4 discusses the empirical methodology. Section 5 presents the results and the final section summarises and concludes the study.

2. Health shocks and top incomes

This chapter is grounded in two main strands of the literature. Firstly, the literature about the impact of health shocks on labour market outcomes. Secondly, the theoretical and empirical literature about the determinants of top incomes, in particular the working super-rich, defined as individuals who receive high earnings due to their employment.

The relationship between health shocks and labour market outcomes is well established in the economic literature, with empirical evidence provided for both developed countries (e.g., García-Gómez and López-Nicolás, 2006) and for developing countries (e.g., Wagstaff 2007, Wagstaff and Lindelow 2014, Mitra *et al.* 2016). However, as anticipated in the introduction, most of the contributions focus on labour participation as the main outcome (e.g., Jones *et al.* 2010, Cai *et al.* 2014) and use data based on elderly workers (e.g., Bound *et al.* 1999; Disney *et al.* 2006; Lee and Kim, 2008; Trevisan and Zantomio, 2016). The main findings show that health shocks have a negative impact on several dimensions, including consumption and income. For example, García-Gómez and López-Nicolás (2006), using Spanish data from the European Community Household Panel, find that a transitory health shock has a negative impact on workers' labour income that ranges between 1400 and 1700 Euros, according to the length of the incapacity period. Halla and Zweimüller (2013) find that Austrian workers who experienced health shocks while commuting incur persistent income losses of about 2-3%, depending on the attachment to labour market. García-Gómez *et al.* (2013), using Dutch hospital and tax register data, find that an acute hospital admission lowers the employment probability by 7% and results in a 5% loss of personal income two years after the shock.

A second strand of literature is mainly based on the "superstars" theories offered in the seminal contributions of Rosen (1981) and Adler (1985), as the basis for disentangling the determinants of super-earnings. In fact, they both argue that superstars arise in markets characterised by imperfect substitution on the demand side and joint consumption on the supply side, which generate a demand concentration towards the better performers (i.e., sportsmen, singers, artists, etc.), who "win and take all". However, while for Rosen (1981) marginal differences in talent are magnified into huge earnings for the most talented due to the convexity of the revenue function, Adler (1985) argues that popularity is the main determinant of superstars' earnings.

From an empirical point of view, a number of researchers have made use of sports data to test these theories. For example, Lucifora and Simmons (2003) investigate "superstar" effects in the wage determination among football players in the Italian Serie A. They find earnings to be highly convex in measures of performance, after controlling for a set of personal characteristics and team fixed effects. Carrieri *et al.* (2017) finds that talent, popularity and agent's bargaining power are all jointly significantly associated to higher wages of Italian football players but their impact is very heterogeneous along the distribution of earnings. Franck and Nuesch (2012), using data from the German Bundesliga, find that both talent and popularity significantly contribute to increasing the market value of superstars. Other studies (e.g., Mullin and Dunn 2002, Treme and Allen 2009, Treme and Allen 2011) focus on American professional sports, finding a positive effect of both measures of performance and media exposure on the entry earnings of baseball (MLB) and basketball (NBA) players.

Building on these strands of literature, this study aims at reconciling the evidence cited above and investigating the impact of a health shock on the labour market outcomes for a specific category of working super-rich: the professional football players of the Italian Serie A.

3. Data and variables

We created an original dataset recording information about wages, performance, injuries and other individual characteristics of the professional players of the Italian Premier League (Serie A), using several data sources. As a starting sample, I analysed data relating to 469 players who have had at least one appearance in the 2013-2014 football season - excluding goalkeepers, following the standard approach of this kind of literature (Lucifora and Simmons, 2003), since their performance is measured differently and they have a dissimilar exposure to the risk of injury compared to the outfield players. The careers of the players were then followed over a 5-season period, from 2010-2011 to 2014-2015. This provides a longitudinal dataset of 1,585 observations. The panel is unbalanced. In fact, the relegation/promotion system between Serie A and Serie B and the transfer market across national and international clubs generates a relatively large turnover of players in the league[21].

Data on players' yearly wage - recorded net of taxes and excluding any performance-related bonus – are taken from the annual report, published at the beginning of each football season by the most influential Italian sport newspaper, *La Gazzetta dello Sport*. Importantly, focusing only on the fixed part of the wage allows for better understanding the effect of the shock on the stock of human capital and reduces concerns about reverse causality, since the dependent variable does not include bonuses that depend on performance. Moreover, to assess the impact of the injury on the renegotiation of the contract, I built a dichotomous variable with value 1 if the contract has been renegotiated at the end of the season with the same club or a new one, in case the player has been traded during the transfer market window, or value 0 in case of no renegotiation.

Information about individual player's characteristics (i.e., birth year, position on the pitch and international appearances) and performance (such as, goals and assists) are extracted from the website *transfermarkt.com*[22]. In addition, in order not to underestimate the performance of midfielders and

[21] There are 20 teams taking part in Serie A. At the end of the season, the last three are in the table are relegated to the second division (Serie B) and replaced by the first three of Serie B.

[22] A German website recording information about football statistics, results, fixtures and news. Data from transfermarkt.com have been used in previous economic studies (i.e., Bryson et al., 2013).

defenders, I collected data about each player's overall performance. These were recorded as the average of the three most read Italian sport newspapers, *La Gazzetta dello Sport, Il Corriere dello Sport* and *Tuttosport*, which rate the player's performance after every match, with a scale that ranges from 0 (poor performance) to 10 (excellent performance). In fact, most of the marks assigned by the journalists range between 4 and 8.

With respect to injuries, I collected data about any injury that occurred in the seasons analysed. In particular, all the injuries were categorised according to the terminology and classification of Mueller-Wohlfahrt *et al.* (2012). Alongside the type of injury, I collected information about the impact of the injury proxied through the days off, identified as the period of time in which the player has not been available for participating in the sports activity of the club (training and matches), and the number of official championship matches missed by the injured player. These data are taken from *footballmarkt.com*. Furthermore, since in the same season a player could incur more than one injury, to build the variable of interest, I consider the number of days off due to either muscular or traumatic injury and I include a dummy variable to control for the re-occurrence of the same injury during the season.

Concerning the clubs, I analysed the annual balance sheets as approved by the directors of the 27 clubs, matched with the players who took part in Serie A for the seasons considered. In particular, information was recorded about five balance sheet items: total wage expense, net sales, earnings before taxes (EBT) and revenues from ticketing and revenues from television rights.

Other variables include a proxy of popularity, measured through the pre-season number of Google search results obtained each year for each player[23] and several other characteristics, related to both clubs and player, which are used as controls in the estimates. These are presented, along with some summary statistics for all the variables, in the next section.

3.1 Descriptive statistics

All the variables included in my dataset, along with their mean values and standard deviations, are presented in Table 1. According to data used in this research, the average annual net wage of a player, in Italian Serie A, amounts to about 840,000 Euros, but with a very large standard deviation (891,000 Euros). Figure 1 shows the non-parametric estimate of the wage distribution for the reference football season 2013-2014. The distribution is positively skewed with a long upper tail. This supports the idea of a restricted number of players who earns huge wages compared to the rest of the distribution and it is consistent with the theories of "superstars" emerging (Rosen 1980, Adler 1985).

[23] Data refers to the same day for each player. They have been collected browsing "name-surname-team" in order to reduce any bias due to homonymy with respect to some surnames, which have higher incidence in Italy.

Indeed, the Gini index within the sampled football players group is 0.47, showing a large degree of inequality even within this group of "privileged" individuals[24].

Concerning the length of the contract, evidence suggests that contracts in the Italian Serie A usually last for 5 years, with the exception of young and old players, that are often characterized by shorter-term arrangements (Carmichael *et al.* 2012). Contracts can be renegotiated even before the expiry date (often to contrast claims by other teams) and in some cases automatic earnings increases are established by the contract according, e.g., to the number of games played during a season. my data show that approximately 65% of players renegotiate the contract each year (56% renegotiate within the same team). This reinforces the focus on a particular type of job market in which the contracts quickly reflect the changes in terms of both productivity (performance) and health shocks.

Concerning injuries, 50.5% of players included in the sample had at least one injury, of which 41.7% were muscle related, while the remaining 58.3% were caused by some kind of traumatic event, such as fractures, cruciate ligament rupture, etc. The re-occurrence rate is 10.4%. In the reference season 2013-2014, on average, a player spent 21 days off activity and missed about 4 official matches. Moreover, information about the exact date of the injury was also collected. Figure A.1 (in the Appendix) shows the percentage distribution of the injuries alongside the different months of the year. Interestingly, the highest percentage of injury was recorded in September (about 15%), which is the first month of official matches after the summer break, followed by the winter months, probably due to worse weather conditions and player fitness. This heterogeneity is controlled for through the inclusion of month fixed effects in some model specifications.

In this chapter, I focus in particular on traumatic injuries since they can be interpreted as an exogenous health shock. This assumes that the probability of such an exogenous shock is not correlated with any observed or unobserved characteristics that are related to the outcomes. Figure 2 shows the non-parametric wage distribution comparing the group of those having at least one traumatic injury during the season and the control group of those who did not. The two distributions are highly similar, showing a long upper tail and an asymmetric distribution, a characteristic already observed above, when analysing the whole sample. However, as reported in Table 2 along with the

[24] Estimates of the Gini index are in line with results found for other soccer leagues. For instance, Franck and Nuesch (2006) find a Gini of 0.56 for the German Bundesliga in the 2004-2005 season (but referring to market values that include also bonuses and potential transfer fees and are then likely to be more dispersed than wages), while Send (2016) finds a Gini of 0.51 considering the earnings of the German Bundesliga players in the 2014-2015 season. A higher inequality emerges instead in the US Major Soccer League: Reilly and Witt (2007) find a Gini of 0.628 in 2007 (0.569 excluding from the computation Beckham who was the most paid player at that time).

relevant percentiles of wage distribution, a large dispersion emerges but the distribution is fairly similar within both groups. In fact, the within-groups inequality is slightly larger among the injured players, with the Gini indices of 0.45 and 0.43, respectively.

Furthermore, to ensure that the two groups do not systematically differ due to pre-injury observable characteristics and to investigate possible concerns of selection bias, I report descriptive statistics for the relevant covariates in Table 3. Remarkably, the mean ages of the two groups do not show a large difference, being 27.9 and 27.3, respectively. The injured players on average earned 95,000 euro more than the controls and played 200 minutes less. However, even if the average wage gap between the two groups seems to be large in absolute terms, it is important to notice that since the study analyses a sample of individuals with very high earnings, the relative size of the gap is less concerning. Indeed, a Student's t-test of the means suggests that these differences are not statistically significant at the 5% level, reinforcing the idea that such a shock may be regarded as random.

Finally, Table A.1 reports the clubs' characteristics regarding some relevant balance sheet items. The clubs have great differences in wage bills with the richest clubs paying almost 7 times more than the bottom ones. Indeed, this inequality is also reflected in the television rights distribution, which is fairly skewed among the clubs. Substantial heterogeneity emerges among Serie A clubs suggesting the need for controlling for team fixed effects in the estimates.

4. Empirical strategy

We aim to obtain causal estimates related to the impact of a health shock on three dimensions: wages, probability of renegotiating the contract and productivity. Firstly, a fixed effects model is used to assess the impact of an injury on the player's wage while controlling for unobserved heterogeneity and an IV strategy is used to handle potential endogeneity concerns and random measurement errors. Secondly, a fixed effects logit model is used to assess the probability of contract renegotiation following the injury.

4.1 Fixed-effects model

We make use of the fixed effects model as principal specification. In fact, after performing the Hausman test, it strongly rejects the hypothesis that the random effects model provides consistent estimates (Woolridge, 2002). The main strength of the FE model is that it takes into account time-invariant observed and unobserved individual characteristics. Controlling for time-invariant unobserved heterogeneity eliminates the effect of some individual peculiarities, such as their attitude to training and genetic or physical characteristics, which might have an effect on both the health shock and the outcome.

Thus, the model to be estimated is the following Mincerian wage equation:

$$\boldsymbol{Log(W)_{it} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 P_{i,t-1} + \gamma S_{i,t-1} + \delta Season_t + u_{it}} \qquad \textbf{(1)}$$

where, the dependent variable is the logarithm of the annual player's wage in season t, net of taxes and bonuses. The coefficient of interest is $\boldsymbol{\gamma}$, where the covariate $\boldsymbol{S_{it-1}}$ represents the incidence of the health shock through the number of days off due to the reported injury. $\boldsymbol{X_{it-1}}$ represents a set of individual time-varying characteristics that are used as controls. All the variables are time lagged to reduce potential reverse causality concerns. Specifically, I include the following controls: age, the square of age, a number of experience related characteristics (e.g., the number of international appearances with either the senior national team or the under 21, a dummy for the team's captain), the total number of minutes played in the previous season. Moreover, I include dummies for the player's position on the pitch (defenders, midfielders and forwards), dummies to distinguish the players nationality (Italian, EU and extra EU) and season dummies. $\boldsymbol{P_{it-1}}$ is a set of covariates which includes proxies of talent and popularity to take into account the main explanations of superstars' earnings (Rosen 1980, Adler 1985). These are goals scored, assists served, ratings given by the newspapers and the number of Google search queries. $\boldsymbol{u_{it}}$ is the composite error term, including both the individual specific and the idiosyncratic error terms.

4.2 Instrumental variable strategy

While the fixed effects model controls for time invariant individual characteristics, unobserved idiosyncratic factors affecting both the health shock and the outcome of interest might be still an issue. Thus, to address concerns about the endogeneity of the health shock and possible measurement errors, I exploit a fixed-effects instrumental variables strategy (FE-IV).

The first stage equation is defined as follows:

$$\boldsymbol{S_{it-1} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 P_{i,t-1} + \gamma z_{i,t-1} + \delta Season_t + u_{it}} \qquad \textbf{(2)}$$

where the instrument **z** is assumed to satisfy ***cov(S,z)≠0*** and ***cov(u, z)=0*** so that it is correlated with the endogenous variable ***S*** and uncorrelated with the error term ***u***. In the FE-IV estimator, the fitted values of $\boldsymbol{\widehat{S}}$ derived from equation (2) are plugged into the original regression equation:

$$\boldsymbol{Log(W)_{it} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 P_{i,t-1} + \gamma \widehat{S}_{i,t-1} + \delta Season_t + u_{it}} \qquad \textbf{(3)}$$

As an instrument I make use of average value of the yellow cards received by the team *j*, excluding those of the player *i*. This excludes the individual's own effort and creates a more pure peer effect, which is correlated with the individual "aggressiveness" put in the game but not with his own wage. Yellow cards have the advantage of measuring the severity of the fouls and should be exogenous being assigned by an external figure, the referee. Indeed, I also might have used information about the yellow cards received by the opponent teams. Yet, the fact that the model takes the football season as the unit of time, implies that in one season every team faced the others twice, thus using the opponent's yellow cards as an instrument would not show enough dispersion.

There are a number of conditions to hold for the instrument to be reliable. Firstly, the instrument has to show some variation over time. This is guaranteed by the fact that, as argued above, the trading of players generates a large turnover of players among teams. The same is also true for the managers. Thus, every season, the players play alongside different teammates and face a different team attitude, according to their team's aims and manager's strategy.

Secondly, the instrument has to be correlated with the potential endogenous variable "days off due to injury". In the framework, yellow cards can be used as a measure of the "aggressiveness" of the team. Thus, it is correlated to the probability of being injured through two different channels. On one hand, a more aggressive team might incur a higher risk due to their attitude towards the game. On the other hand, due to a mechanism of reciprocity, the opponent team might play in a tougher way in response, increasing the number of foul plays and, consequently, the probability of injuries.

Lastly, the instrument, conditional on the health shock, should not affect the individual wage. With respect to this, I make use of the average yellow cards of the player's *i* teammates, which may affect his game style through a "peer effect", but it is unlikely to have any correlation with his own wage.

4.3 Binary choice models

One main advantage of using sports data for labour market analysis is the availability of information about the contract between the player and the team, which can be renegotiated at any point before it expires. There are two main reasons for renegotiating the player's contract. First, if the contract is close to expiry, the club might not want to risk losing a player without any monetary compensation. Second, the player may be traded to another club during one of the two transfer market periods, thus entailing the signing of a new contract. To assess the impact of the health shock on the probability of renegotiating the contract at the end of the season, I make use of two different binary choice models.

Firstly, a linear probability model (LPM) is used to exploit the longitudinal nature of the data and the IV strategy, with fixed effects to control for individual time-invariant characteristics. I estimate the following equation:

$$\boldsymbol{d_{it} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 P_{i,t-1} + \gamma S_{i,t-1} + \delta Season_t + u_{it}} \qquad \textbf{(4)}$$

where $\boldsymbol{d_{it}}$ is a dichotomous variable assuming either value 1, if the contract between the player and the club has been renegotiated at the end of the season, or value 0 if it had not. The covariate $\boldsymbol{S_{it-1}}$ represents the incidence of the health shock through the number of days off due to the reported injury. The remaining set of covariates is the same as described in the equation (1). I use standard heteroscedasticity-robust standard errors and *t*-statistics to deal with heteroscedasticity. Furthermore, to reduce any concern about the endogeneity of the health shock, I implement the IV strategy as explained above, using the 2SLS estimator and instrumenting the injury through the average value of the team's yellow cards, excluding those of the individual *i*.

Secondly, I use a fixed-effect logit model. The FE logit allows for both the individual time invariant characteristics and the dichotomous nature of the dependent variable. Compared to the LPM, the FE logit gives a response probability that ranges from 0 to 1. However, one drawback of the FE logit models is that the interpretation of the results is somewhat cumbersome, due to the problem of computing predicted probabilities of the outcome and the marginal or discrete effects when the fixed effect is unknown. To overcome this issue, it is common practice in the empirical literature to interpret the effect in terms of odds ratio or on the conditional probability (Cameron and Trivedi, 2010). Indeed, both these methods are less intuitive when it comes to the economic interpretation. In this chapter, I interpret the FE logit model coefficients in terms of odds ratios.

5. Results

5.1 Strength of the instrument

Before presenting the main results, I assess the validity of the IV approach. Table A.2 in the Appendix presents the first-stage regressions on the full set of covariates, for the different specifications. The instrument is strongly correlated with the health shock measure. Thus, results indicate that playing for a more aggressive team is associated with a higher number of days off, significant at 1% level, confirming the intuition behind the relevance of the instrument.

Furthermore, at the bottom of the Table A.2, I report the F-test results of whether the excluded instrument significantly differs from zero. The F-statistic is always above 10, which is generally adopted as "rule of the thumb" for the minimum threshold to reject the null hypothesis of a weak instrument (Staiger and Stock, 1997). In addition, Stock and Yogo (2005) provided a more formal threshold (16.38) to test the weakness of the instrument. My instrument's F-test is also above this critical value, ranging between 18.34 and 20.45 across the different specifications.

5.2 Effects of health shock on wages

Tables 4-6 show the results of the fixed effects model described in Equation (1), with the logarithm of wage as the dependent variable. Table 4 shows the results for the full sample, Table 5 controls for muscular injuries and Table 6 shows results for a subsample in which muscular injuries are dropped. Column 1 of each table reports the results of the FE model, without team and month fixed effects. Column 2-3 include results with either team or month fixed effects, Column 4's results account for both. Moreover, to make the economic interpretation of the coefficient of interest easier and comparable, I account for 30 days variation in the number of days off due to injury. Other covariate coefficients are presented according to a change of one standard deviation, as reported in the summary statistics (Table 1). Since the dependent variable is in log format, the estimated coefficients report a percentage variation in the net annual wage associated with a one standard deviation increase of the independent variable.

Table 4 shows a negative relationship between days off due to injury and the net wage. In particular, one-month of injury is associated with a reduction of about 2.5% in the wage of the following season. The size and significance of the coefficient remains the same when controlling for team fixed effects and it slightly increases where month fixed effects are controlled for.

With respect to the other covariates, Table 4 shows that both indicators of performance and popularity are significant and have a positive effect on the annual net wage. In particular, in the baseline model (Column 1), a one-S.D. increase in goals and popularity affects the following season's wage by about 4.9% and 3.2%, respectively. Coefficients remain almost unaltered when accounting for team and month fixed effects.

These results are robust to different model specifications, when controlling for muscular injuries (Table 5) or dropping them from the sample (Table 6). Indeed, in Table 5, the coefficient of days off due to muscular injuries is not statistically significant. This result supports the intuition that the health shock effect on wages is due to those injuries which have an underlying random component, such as the traumatic injury types.

Interestingly, it is possible to observe a larger spread between coefficients when month fixed effect are included. These results indicate that, during the wage bargaining process, variation across the season should be accounted for. In particular, some months might be associated with worse weather conditions or with a larger number of matches played due to the calendar of fixtures, which exposes players to higher stress in some periods of the season.

The results for the FE-IV estimation on the full set of controls are shown in Table 7. Making use of the teammates' yellow cards as an instrument, results in a significant increase in the coefficient of the days off variable. A one-month injury has a more negative effect on the wage, reducing it by about 12%; the coefficient is statistically significant at 5% level.

With respect to the other covariates, a one-S.D. increase in goals and popularity leads to a positive effect on the annual net wage of about 5.4% and 3.4%, respectively. The effect is slightly higher than in the FE model and the coefficients concerning goals are still significant at the 5% level while the for popularity they drop to the 10% level, showing less spread among the different model specifications, when team and/or month fixed effect are controlled for.

Moreover, results show a non-linear relationship between age and the wage. In particular, wage increases with age due to the cumulated experience and knowledge of the game until a turning point in which these time varying personal characteristics are overcome by the progressive loss of physical skills. I estimated this turning point to be at around 32 years old, higher than in the early-2000 evidence (Lucifora and Simmons, 2003) and consistent with the evidence of the elongation of players' careers observed in the last 15 years.

Table 8 reports estimates that exclude the performance-related covariates. This further specification disentangles the effect of the injury on the wage between being a direct effect and an indirect one, which is mediated by the performance. It is interesting to notice that when the model does not control for performance, the coefficient of the days off increases by around 1%, across the different specifications. Thus, it is possible to argue that the effect of the injury is mediated through performance, due to reduced physical skills, by only a small amount. Indeed, these results can be seen in a framework of human capital depreciation as showing that during the bargaining process the club is willing to offer a lower amount to the player for precautionary reasons rather than because it is concerned by the negative effect of the injury on the individual's productivity.

5.3 Effects of health shocks on renegotiation

Tables 9 and 10 show the results of the binary choice models of wage renegotiation with and without controlling for performance, respectively. Table 10 thus includes in the sample also those individuals whose performance was not measurable due to the fact that they did not play any official match, because of season-long injuries. Columns 1-4 of both tables report estimates of the linear probability model described in Equation (4) while columns 5-8 report results of the fixed-effect logit model.

Table 9 shows that health shocks are positively associated with the probability of renegotiating the contract at the end of the season. The coefficient is anyway not statistically significant at conventional levels. However, when including in the sample also long-season injured, as in Table 10, injury effect turns to be negative and statistically significant. Specifically, a 30-day injury is found to increase the probability of renegotiating the contract by approximately 9.8%. Results show similar patterns when considering the FE logit model, in terms of odds ratio.

Thus, from these results an interesting dynamic emerges. Injury is found to have a significant effect on the probability of renegotiation of the contract. This effect is even more prominent if the injury resulted in worse performances or persisted throughout the entire season, preventing the player from taking part in any official match. It follows that a player, either due to the reduced performance or because he fails to demonstrate a degree of reliability measured in terms of the number of games played, is found to accept a less-paid contract by the same club or to be compelled to migrate to another club at the end of the season and thus to renegotiate his contract.

Concerning the other covariates, every additional 90 minutes played increase the probability of renegotiating the contract by about 0.85 points. Essentially, the clubs are interested in extending the contracts of those players who guarantee a higher reliability in terms of both physical and tactical integrity and experience, cumulated through the larger number of matches played.

Interestingly, a 1 S.D. increase in popularity is negatively associated with the contract renegotiation. This result is indicative of a well-known dynamic in the football market. In fact, when a player's popularity increases they have an increased incentive to delay the contract's renegotiation with their own club with the aim of arriving as close as possible to the expiration date in order to threaten the club into offering them a higher wage so as not to lose the player without any monetary compensation, at the end of the contract. Furthermore, a more popular player represents for the club an important asset in terms of merchandising and image rights. Hence, the bargaining process becomes much more complex and the timing plays a central role.

6. Robustness checks

To check the robustness of these findings, I perform a number of sensitivity analyses. Firstly, an alternative measure of health shock, namely the total number of official league matches in which the player was reported as "not available due to injury" is used to check whether the results are confirmed. In fact, it may be argued that the maximum number of days off is specific to the form of injury and not influential in given periods of the year if it does not overlap with any official match. Nevertheless, the channel through which the club decides to offer a lower wage might be based on the actual rate of participation of the player in official matches. The estimates, reported in Table 11, show that the results are robust based on this alternative health shock measure. In fact, every additional match missed negatively affects the wage for the following season by approximately 2.8%. This result is consistent for both magnitude and sign with the main results. In fact, in one month, on average, a club has scheduled about 5 Serie A matches, making this result proportional to the baseline one, in which 30-day variations in the number of days off were considered. Furthermore, this alternative model confirms the strength of the instrument, whose F-test ranges between 16.56 and 19.08, across the different specifications, as reported in the first stage F-test at the bottom of Table 11.

Secondly, I test whether the dynamics of the model are correctly specified. In fact, under multi-year contracts, the effect on the wage might have been determined by a health shock which occurred in a previous season compared to the one of the baseline model (i.e., t-2 or t-3). Thus, I run the model including the lagged values of days off, up to three seasons before. Table 12 shows that even though the number of days off in earlier seasons is negatively associated with the net wage, the coefficients are not statistically significant at conventional levels, reinforcing the idea that the negative effect is caused by the injury occurred just in the previous season with respect to the eventual contract renegotiation.

Thirdly, to check the robustness of the hypothesis about the existence of a difference between the direct effect of the health shock on the outcome and the indirect one, mediated through performance, I estimate performance measures that are not affected by injury. In particular, I proxy the non-injury affected performance by the OLS residuals of the regression of the performance's measures (i.e., grade, goal and assist) on the number of days off, while controlling for age and seasonal fixed effects. Then, I include the injury-purged variables as covariates in the FE-IV estimation. The results, reported in Table 13, show that the magnitude and the sign of the health shock index remains unaltered compared to the baseline model. Interestingly, in this specification, the grade of the overall performance becomes statistically significant: a 1 S.D. increase in the injury-purged performance positively affects the net wage of the following year by approximately 4.7%. This strengthens the evidence in support of the existence of both a direct and indirect effect of the health shock. Furthermore, the first stage regressions reported in Table 13 show a slight improvement of the power of the instrument in this specification and the F-tests range between 19.77 and 21.98.

7. Conclusions

This chapter exploits data from professional football to estimate the effect of health shock on labour market outcomes. Using a unique longitudinal dataset created by recording information about several characteristics of the universe of football players of the Italian Serie A followed from 2009 to 2014, my analysis makes a number of substantial contributions to the extensive literature on the labour market consequences of health shocks.

Firstly, I report evidence on the effect of the shock on the outcomes of a significant share of the working super rich. Working super rich phenomenon started to rise considerably from 1970's in all developed countries attracting the interest of many scholars (Atkinson *et al.*, 2011). Previous analyses focused on several dimensions of this phenomenon, including the determinants of the high earnings and the effects on overall inequality. However, to the best of my knowledge, no previous studies focused on the effect of a health shock on their earnings. Health might be instead a key dimension to better understand the dynamics of contemporary society's extreme inequalities. Secondly, I exploit the exogenous nature of the traumatic injuries and the availability of panel data to retrieve causal estimates of the effect of the health shock on labour market outcomes. Moreover, I employ an instrumental variable strategy which use the average number of yellow cards received by the team as an instrument to reduce any further endogeneity concerns. This is a rare feature of studies analysing the relationship between health shocks and outcomes on cross-sectional survey data. In these settings, it is generally hard to deal with the endogenous relationship between health shocks and labour market outcomes. Lastly, professional sports data provides detailed measures of worker's productivity. This is useful in order to disentangle the main channels through which health shock affects labour market outcomes in my setting.

The main findings of this chapter can be summarised as follows. Firstly, I find that health shocks have a strong negative effect on the wages of the professional football players of the Italian Serie A. In particular, having a 30-day injury reduces the wage of the following year by approximately 12%. Secondly, I find that only a residual part of the negative effect could be explained through the reduced performance of the players after the injury. In fact, results suggest that the largest part of the coefficient is explained by a direct effect of the shock on the outcome. This result can be explained in a framework of human capital depreciation. The club has an incentive to offer a lower wage for precautionary reasons, supposing that players who experienced a severe injury might incur similar injuries in the future. Thus, the club insures itself against this risk by reducing the fixed share of the wage, which is independent by performances.

Furthermore, I find that the effect of the injury on the probability of renegotiating the contract, at the end of the season, increases with its severity. In fact, different model specifications show that having a season-long injury increase the probability of renegotiating the contract with the same club or another one by about 10%. This result suggests that the effect is explained by both the detrimental effect of the injury on performance and the precluded participation of the player in official matches. Finally, in line with the theoretical and empirical literature about super-earnings, If ind that performance and popularity play an important role in explaining the wages of the "working super-rich".

Overall, these findings suggest that also working super-rich are negatively affected by health shocks in terms of reduced earnings but with interesting peculiarities with respect to the "ordinary workers". First, the magnitude of the impact of a health shock on working super-rich's income seems to be significantly larger than the effect on other workers. For instance, Halla and Zweimüller (2013) find an income loss of about 2-3% among workers incurring in accidents when commuting in Austria, while García-Gómez *et al.* (2013) report a 5% loss of personal income following acute hospital admission in the Netherlands. I find instead a wage penalization of 12% for an "only" 30-days injury. Thus, the exposure to the health shock seems to have more dramatic economic consequences for super-rich workers and the high salaries they receive are likely to incorporate also a premium against the high costs associated to the health shocks. The average short length of the contract represents an additional element of uncertainty in the career of these workers and may contribute to further exacerbate the negative consequences of the health shocks. Secondly, the precautionary reasons underlying the wage penalization represents a substantial difference with respect to the effect of health shocks in other parts of the income distribution. This might be partly due to the activity of the football player largely based on physical skills that might raise employee's concerns around his future productivity. My results are then likely to apply also other categories of super-rich whose rewards reflect competitive performance and popularity with the public. On the other hand, the large effect of health shock on these categories might be also due to the characteristics of the markets in which these workers are generally involved. These are characterised by frequent contract renegotiations and high job turnover.

Further research might investigate whether the patterns observed in my setting apply also other categories of super-rich workers involved in less-physically demanding jobs, such as the CEOs. Similarly, an analysis on data from other professional sports or other football leagues might be useful to understand the degree of generalizability of my results to other categories of professional athletes.

References

Adler, M. (1985). Stardom and talent. *The American Economic Review*, *75*(1), 208-212.

Alvaredo, F., Atkinson, A. B., Piketty, T., & Saez, E. (2013). The top 1 percent in international and historical perspective. *The Journal of Economic Perspectives*, *27*(3), 3-20.

Atkinson, A. B., Piketty, T., & Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, *49*(1), 3-71.

Bound, J., Schoenbaum, M., Stinebrickner, T. R., & Waidmann, T. (1999). The dynamic effects of health on the labor force transitions of older workers. *Labour Economics*, *6*(2), 179-202.

Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: footedness and player remuneration in European soccer. *Journal of Sports Economics*, *14*(6), 606-628.

Cai, L. (2010). The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model. *Labour Economics*, *17*(1), 77-90.

Cai, L., Mavromaras, K., & Oguzoglu, U. (2014). The effects of health status and health shocks on hours worked. *Health Economics*, *23*(5), 516-528.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata* (Vol. 2). College Station, TX: Stata press.

Carmichael, F., Rossi, G., Simmons, R. (2012). Contract Duration and Player Performance in Italian Football, *4th European Conference in Sport Economics / XIV IASE*, Birkbeck University of London.

Carrieri, V., Principe, F., & Raitano, M. (2017). What makes you" super-rich"? New evidence from an analysis of football players' earnings *Ruhr Economic Papers,* no. 681.

Charles, K. K. (2003). The longitudinal structure of earnings losses among work-limited disabled workers. *Journal of human Resources*, *38*(3), 618-646.

Deaton, A. (2003). Health, inequality, and economic development. *Journal of economic literature*, *41*(1), 113-158.

Deloitte (2017). Football money league. 20th Edition. *Sports Business Group.*

Disney, R., Emmerson, C., & Wakefield, M. (2006). Ill health and retirement in Britain: A panel data-based analysis. *Journal of Health Economics*, *25*(4), 621-649.

Franck, E., Nüesch, S. (2006), Talent, Past Consumption and/or Popularity - Are German Soccer Celebrities Rosen or Adler Stars?, *Working Papers 0043, University of Zurich, Institute for Strategy and Business Economics* (ISU).

Franck, E., & Nüesch, S. (2012). Talent and/or popularity: what does it take to be a superstar?. *Economic Inquiry*, *50*(1), 202-216.

Franzini, M., Granaglia, E., & Raitano, M. (2016). *Extreme Inequalities in Contemporary Capitalism: ShouldIbe Concerned about the Rich?*. Springer.

García-Gómez, P., & López Nicolás, Á. (2006). Health shocks, employment and income in the Spanish labour market. *Health Economics*, *15*(9), 997-1009.

García-Gómez, P., Van Kippersluis, H., O'Donnell, O., & Van Doorslaer, E. (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources*, *48*(4), 873-909.

Halla, M., & Zweimüller, M. (2013). The effect of health on earnings: Quasi-experimental evidence from commuting accidents. *Labour Economics*, *24*, 23-38.

Hanson, A., Jolly, N. A., & Peterson, J. (2017). Safety regulation in professional football: empirical evidence of intended and unintended consequences. *Journal of Health Economics*, *53*, 87-99.

Jones, A. M., Rice, N., & Roberts, J. (2010). Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS. *Economic Modelling*, *27*(4), 866-880.

Kahn, L. M. (2000). The sports business as a labor market laboratory. *The Journal of Economic Perspectives*, *14*(3), 75-94.

Lee, J., & Kim, H. (2008). A longitudinal analysis of the impact of health shocks on the wealth of elders. *Journal of Population Economics*, *21*(1), 217-230.

Leigh, A. (2007). How closely do top income shares track other measures of inequality?. *The Economic Journal*, *117*(524).

Lindeboom, M., & Kerkhofs, M. (2009). Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work. *Journal of Applied Econometrics*, *24*(6), 1024-1046.

Lucifora, C., & Simmons, R. (2003). Superstar effects in sport: Evidence from Italian soccer. *Journal of Sports Economics*, *4*(1), 35-55.

Mitra, S., Palmer, M., Mont, D., & Groce, N. (2016). Can households cope with health shocks in Vietnam?. *Health Economics*, *25*(7), 888-907.

Mueller-Wohlfahrt, H. W., Haensel, L., Mithoefer, K., Ekstrand, J., English, B., McNally, S., ... & Blottner, D. (2012). Terminology and classification of muscle injuries in sport: a consensus statement. *British Journal of Sports Medicine*, bjsports-2012.

Mullin, C. J., & Dunn, L. F. (2002). Using baseball card prices to measure star quality and monopsony. *Economic Inquiry*, *40*(4), 620-632.

Reilly, B., Witt, R., (2007). The Determinants of Base Pay and the Role of Race in Major League Soccer: Evidence from the 2007 League Season, *School of Economics Discussion Papers 1907*, School of Economics, University of Surrey.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, *71*(5), 845-858.

Send. J. (2016), Football and Money: Income Inequality in the German Bundesliga, https://the10thmanblog.wordpress.com/

Staiger, D., & Stock, J. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica, 65*(3), 557-586.

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D.W.K., & Stock, J., (Eds.), Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg, 80-108. *New York: Cambridge University Press.*

Stoecker, C., Sanders, N. J., & Barreca, A. (2016). Success is something to sneeze at: Influenza mortality in cities that participate in the Super Bowl. *American Journal of Health Economics*, 2 (1), 125-143.

Trevisan, E., & Zantomio, F. (2016). The impact of acute health shocks on the labour supply of older workers: Evidence from sixteen European countries. *Labour Economics*, *43*, 171-185.

Treme, J., & Allen, S. K. (2009). Widely received: Payoffs to player attributes in the NFL. *Economics Bulletin*, *29*(3), 1631-1643.

Treme, J., & Allen, S. K. (2011). Press pass: Payoffs to media exposure among National Football League (NFL) wide receivers. *Journal of Sports Economics*, *12*(3), 370-390.

Wagstaff, A. (2007). The economic consequences of health shocks: evidence from Vietnam. *Journal of Health Economics*, *26*(1), 82-100.

Wagstaff, A., & Lindelow, M. (2014). Are health shocks different? Evidence from a multishock survey in Laos. *Health Economics*, *23*(6), 706-718.

Wooldridge, J. M. (2002) Econometric Analysis of Cross Section and Panel Data. *The MIT Press.*

Tables and Figures

Table 1. Sample characteristics

Variable	*Description*	*Mean (St. dev.)*
Dependent variable		
Wage	Net earnings (pre-season values) in thousands/€	875.1 (911.7)
Log wage	Log of net earnings (pre-season values)	6.38 (0.87)
Renegotiation	Dummy for contract's renegotiation	
Individual controls		
Age	Age (years)	26.6 (4.2)
Age square	Age squared	725.8 (226.7)
Position	Dummies for defenders (40.2%), midfielder (39.9%) and forward (19.9%)	
Captain	Dummy for the team's captain	0.034 (0.181)
Minutes played	Minutes played during the season	1352.8 (1068.3)
Total international caps	Number of caps with the national team up to 2014-2015	15.75 (25.10)
Total Under-21 caps	Number of caps with the U21 national team up to 2014-2015	5.85 (8.68)
International caps	Number of caps with the national team during the season	2.01 (4.65)
Under-21 caps	Number of caps with the U21 national team during the season	0.37 (1.59)
Player's performance		
Grade	Mean grade by newspapers during the season	5.77 (0.41)
Goal	Goals scored during the season	1.93 (3.50)
Assist	Assists served during the season	1.28 (2.04)
Index of popularity		
Popularity	Google search results (million)	4.21 (9.37)
Index of health shock		
Injury	Dummies for the kind of injury	
Days-off traumatic	Number of days off due to traumatic injury	11.08 (38.77)
Days-off muscular	Number of days off due to muscular injury	4.77 (12.90)
Re-occurrence	Dummy for the re-occurrence of the same kind of injury	4.79 (7.35)
Matches off	Total number of matches missed due to any injury	4.04 (7.12)
Instrumental variable	Player/Team	
Yellow cards	Number of teammates' avg. yellow cards	1.91 (0.48)

Table 2. Wage distribution: percentiles and Gini Index

	Total	Treatment	Control
Mean	985.9	1050.9	955.8
Standard Deviation	941.3	1025.1	898.9
Minimum	30	30	30
p10	300	300	300
p25	400	400	400
p50	650	700	600
p75	1200	1200	1100
p90	2100	2400	2100
p99	4500	4900	4200
Maximum	6500	5500	6500
Gini Index	0.474	0.453	0.435

Table 3. Within group characteristics

	Treatments				Controls				Mean difference	p-value[1]
	Mean	Standard Dev.	Min	Max	Mean	Standard Dev.	Min	Max		
Wage	1050.9	1025.1	30	5500	955.7	898.9	30	6500	95.2	0.11
Caps	22.49	8.95	1	38	24.5	9.51	0	38	2.01	0.005
Minutes	1651.9	814.0	8	3230	1811.6	904.9	0	3643	159.6	0.004
Age	27.9	4.1	19	40	27.3	3.8	18	39	0.6	0.01
Goal	2.56	4.14	0	28	2.52	3.68	0	29	0.04	0.84
Assist	1.72	2.29	0	12	1.67	2.15	0	14	0.05	0.72
Grade	5.76	0.43	3.33	6.75	5.78	0.39	3.67	6.7	0.02	0.33
Popularity	0.45	0.89	0.003	9.20	0.39	0.87	0.002	9.66	0.06	0.28

[1] p-value for the Student t-test of means comparison. H_0 diff=0; H_a diff≠0

Table 4. Effects of health shock and other covariates on annual net (log)wage. Fixed effects estimates.

	(1)	(2)	(3)	(4)
Age	0.8971***	0.9031***	0.8796***	0.8856***
	0.0639	*0.0640*	*0.0640*	*0.0641*
Age sq.	-0.0150***	-0.0151***	-0.0148***	-0.0149***
	0.0011	*0.0011*	*0.0011*	*0.0011*
Goal	0.0491***	0.0486**	0.0507***	0.0502***
	0.0190	*0.0190*	*0.0189*	*0.0189*
Assist	-0.0024	-0.0031	-0.0045	-0.0053
	0.0144	*0.0144*	*0.0144*	*0.0144*
Grade	0.0181	0.0188	0.0176	0.0184
	0.0149	*0.0149*	*0.0148*	*0.0148*
Popularity	0.0321**	0.0330**	0.0315**	0.0325**
	0.0160	*0.0160*	*0.0159*	*0.0159*
Days-Off traum.	-0.0250**	-0.0250**	-0.0283**	-0.0284**
	0.0118	*0.0118*	*0.0118*	*0.0118*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1197	1197	1197	1197

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Robust Standard Errors in *italics*.

Table 5. Effects of health shock and other covariates on annual net (log)wage. Fixed effects estimates. Days off due to muscular injury as control.

	(1)	(2)	(3)	(4)
Age	0.8924***	0.8984***	0.8786***	0.8847***
	0.0642	*0.0643*	*0.0642*	*0.0643*
Age sq.	-0.0149***	-0.0150***	-0.0148***	-0.0149***
	0.0011	*0.0011*	*0.0011*	*0.0011*
Goal	0.0487**	0.0482**	0.0505***	0.0501***
	0.0190	*0.0190*	*0.0189*	*0.0189*
Assist	-0.0029	-0.0036	-0.0046	-0.0054
	0.0145	*0.0145*	*0.0144*	*0.0144*
Grade	0.0175	0.0183	0.0175	0.0182
	0.0149	*0.0149*	*0.0149*	*0.0149*
Pop	0.0321**	0.0331**	0.0315**	0.0325**
	0.0160	*0.0160*	*0.0159*	*0.0159*
Days-Off traum.	-0.0230*	-0.0230*	-0.0276**	-0.0277**
	0.0120	*0.0120*	*0.0121*	*0.0121*
Days-Off musc.	0.0007	0.0007	0.0002	0.0002
	0.0008	*0.0008*	*0.0008*	*0.0008*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1197	1197	1197	1197

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Robust Standard Errors in *italics*.

Table 6. Effects of health shock and other covariates on annual net (log)wage. Fixed effects estimates. Days off due to muscular injury dropped.

	(1)	(2)	(3)	(4)
Age	0.8051***	0.8059***	0.8000***	0.8015***
	0.0761	*0.0765*	*0.0760*	*0.0764*
Age sq.	-0.0134***	-0.0134***	-0.0134***	-0.0134***
	0.0013	*0.0013*	*0.0013*	*0.0013*
Goal	0.0586**	0.0586**	0.0572**	0.0571**
	0.0248	*0.0249*	*0.0248*	*0.0248*
Assist	0.0097	0.0096	0.0070	0.0068
	0.0189	*0.0190*	*0.0189*	*0.0190*
Grade	0.0397**	0.0398**	0.0401**	0.0403**
	0.0191	*0.0192*	*0.0191*	*0.0191*
Popularity	0.0329	0.0331	0.0321	0.0324
	0.0204	*0.0205*	*0.0204*	*0.0205*
Days-off traum.	-0.0217*	-0.0218*	-0.0281**	-0.0283**
	0.0132	*0.0132*	*0.0136*	*0.0137*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	848	848	848	848

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Robust Standard Errors in *italics*.

Table 7. Effects of health shock and other covariates on annual net (log)wage.
Fixed effects-IV estimates.

	(1)	(2)	(3)	(4)
Days-off	-0.1211**	-0.1204**	-0.1258**	-0.1250**
	0.0563	*0.0558*	*0.0598*	*0.0592*
Age	0.9795***	0.9861***	0.9776***	0.9839***
	0.0856	*0.0857*	*0.0859*	*0.0859*
Age sq.	-0.0161***	-0.0161***	-0.0160***	-0.0161***
	0.0014	*0.0014*	*0.0014*	*0.0014*
Goal	0.0538**	0.0540**	0.0533**	0.0535**
	0.0214	*0.0214*	*0.0216*	*0.0216*
Assist	0.0043	0.0038	0.0033	0.0027
	0.0167	*0.0167*	*0.0169*	*0.0168*
Grade	0.0276	0.0286*	0.0282	0.0293*
	0.0169	*0.0169*	*0.0172*	*0.0172*
Popularity	0.0345*	0.0353*	0.0351*	0.0359*
	0.0183	*0.0183*	*0.0185*	*0.0185*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1099	1099	1099	1099
First Stage F-test	*20.23*	*20.45*	*18.34*	*18.62*

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Standard Errors in *italics*.

Table 8. Effects of health shock and other covariates on annual net (log)wage.
Fixed effects-IV results. Performance not controlled for.

	(1)	(2)	(3)	(4)
Days-off	-0.1303**	-0.1309**	-0.1362**	-0.1367**
	0.0581	*0.0578*	*0.0620*	*0.0615*
Age	1.0168***	1.0246***	1.0141***	1.0217***
	0.0882	*0.0884*	*0.0885*	*0.0887*
Age sq.	-0.0167***	-0.0168***	-0.0167***	-0.0168***
	0.0014	*0.0014*	*0.0015*	*0.0015*
Popularity	0.0334*	0.0343*	0.0341*	0.0350*
	0.0186	*0.0186*	*0.0188*	*0.0188*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1099	1099	1099	1099
First Stage F-test	*19.67*	*19.80*	*17.74*	*17.97*

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Standard Errors in *italics*.

Table 9. Effects of health shock and other covariates on renegotiation.
Binary choice models estimates.

	Linear probability model				Fixed effects logit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Days-off	0.079	0.079	0.054	0.057	1.082	1.083	1.056	1.059
	0.073	*0.073*	*0.075*	*0.075*	*1.08*	*1.09*	*0.72*	*0.76*
Age	0.052	0.046	0.042	0.036	1.053	1.047	1.043	1.036
	0.111	*0.103*	*0.099*	*0.100*	*0.47*	*0.45*	*0.42*	*0.36*
Age sq.	0.001	0.002	0.001	0.002	1.002	1.002	1.002	1.002
	0.001	*0.002*	*0.002*	*0.002*	*0.86)*	*0.97*	*1.01*	*1.05*
Caps	0.008*	0.009*	0.008*	0.008*	1.008*	1.009*	1.008*	1.008*
	0.005	*0.005*	*0.005*	*0.005*	*1.80*	*1.76*	*1.71*	*1.76*
Goal	-0.037	-0.037	-0.039	-0.039	0.963	0.963	0.961	0.961
	0.028	*0.027*	*0.026*	*0.026*	*-1.34*	*-1.41*	*-1.53*	*-1.52*
Assist	0.002	0.002	-0.003	-0.003	1.002	1.002	0.997	0.997
	0.022	*0.022*	*0.022*	*0.022*	*0.09*	*0.10*	*-0.15*	*-0.13*
Grade	0.020	0.020	0.024	0.023	1.021	1.020	1.024	1.023
	0.022	*0.021*	*0.021*	*0.021*	*0.93*	*0.93*	*1.13*	*1.09*
Pop.	-0.040*	-0.041*	-0.037*	-0.038*	0.960*	0.960*	0.963*	0.963*
	0.024	*0.021*	*0.020*	*0.020*	*-1.70*	*-1.94*	*-1.85*	*-1.87*
Team FE	No	Yes	No	Yes	No	Yes	No	Yes
Month FE	No	No	Yes	Yes	No	No	Yes	Yes
N	1099	1099	1099	1099	1099	1099	1099	1099

***, **, * indicate significance at 1%, 5% and 10%, respectively
[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Robust Standard Errors in *italics*.

Table 10. Effects of health shock and other covariates on renegotiation.
Binary choice models estimates. Performance not controlled for.

	Linear probability model				Fixed effects logit			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Days-off	0.098**	0.104**	0.069	0.075	1.103*	1.109*	1.072	1.079
	0.046	*0.047*	*0.049*	*0.050*	*2.13*	*2.25*	*1.40*	*1.50*
Age	0.200**	0.189**	0.189**	0.179**	1.221*	1.208*	1.209*	1.196*
	0.083	*0.079*	*0.074*	*0.076*	*2.42*	*2.38*	*2.55*	*2.36*
Age sq.	-0.001	-0.001	-0.001	-0.001	0.999	0.999	0.999	0.999
	0.001	*0.001*	*0.001*	*0.001*	*-0.91*	*-0.84*	*-0.93*	*-0.80*
Caps	0.010***	0.010***	0.009***	0.010***	1.010***	1.010***	1.010***	1.010***
	0.002	*0.001*	*0.001*	*0.002*	*5.97*	*6.15*	*5.91*	*5.95*
Pop.	-0.031*	-0.032**	-0.028**	-0.029**	0.969*	0.969*	0.972*	0.972*
	0.017	*0.014*	*0.014*	*0.014*	*-1.86*	*-2.18*	*-2.01*	*-2.05*
TeamFE	No	Yes	No	Yes	No	Yes	No	No
Month Fe	No	No	Yes	Yes	No	No	Yes	No
N	1527	1527	1527	1527	1527	1527	1527	1527

***, **, * indicate significance at 1%, 5% and 10%, respectively
[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Robust Standard Errors in *italics*.

Table 11. Robustness check: effects of health shock and other covariates on annual net (log)wage. Alternative health shock's measure. Fixed effects-IV estimates.

	(1)	(2)	(3)	(4)
Matches-off	-0.0282**	-0.0294**	-0.0289**	-0.0301**
	0.0134	*0.0141*	*0.0140*	*0.0147*
Age	1.0003***	1.0117***	0.9996***	1.0109***
	0.0930	*0.0956*	*0.0933*	*0.0960*
Age sq.	-0.0162***	-0.0164***	-0.0162***	-0.0163***
	0.0015	*0.0015*	*0.0015*	*0.0015*
Goal	0.0472**	0.0467**	0.0467**	0.0462**
	0.0220	*0.0223*	*0.0222*	*0.0225*
Assist	0.0098	0.0096	0.0093	0.0090
	0.0177	*0.0179*	*0.0177*	*0.0179*
Grade	0.0355*	0.0369**	0.0361*	0.0376**
	0.0184	*0.0187*	*0.0187*	*0.0190*
Popularity	0.0313*	0.0322*	0.0316*	0.0326*
	0.0187	*0.0189*	*0.0188*	*0.0190*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1099	1099	1099	1099
First Stage F-test	*19.08*	*17.59*	*17.72*	*16.56*

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Standard Errors in *italics*.

Table 12. Robustness check: effects of health shock and other covariates on annual net (log)wage. Lagged health shock's values included. Fixed effects estimates.

	(1)	(2)	(3)	(4)
Age	0.9356***	0.9371***	0.9253***	0.9280***
	0.1319	*0.1313*	*0.1335*	*0.1325*
Age sq.	-0.0155***	-0.0155***	-0.0153***	-0.0154***
	0.0022	*0.0022*	*0.0022*	*0.0022*
Goal	0.0451*	0.0453*	0.0449*	0.0454*
	0.0256	*0.0256*	*0.0251*	*0.0251*
Assist	0.0266	0.0265	0.0252	0.0250
	0.0172	*0.0170*	*0.0171*	*0.0170*
Grade	0.0446**	0.0446**	0.0449**	0.0449**
	0.0206	*0.0206*	*0.0204*	*0.0204*
Popularity	0.0583*	0.0584*	0.0565*	0.0567*
	0.0336	*0.0336*	*0.0330*	*0.0331*
Days-off	-0.0333*	-0.0331*	-0.0348*	-0.0346*
	0.0186	*0.0185*	*0.0187*	*0.0186*
Days-off$_{t-2}$	-0.0220	-0.0220	-0.0201	-0.0201
	0.0215	*0.0215*	*0.0215*	*0.0213*
Days-off$_{t-3}$	-0.0365	-0.0364	-0.0351	-0.0349
	0.0242	*0.0242*	*0.0241*	*0.0241*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	664	664	664	664

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [b] Standard Errors in *italics*.

Table 13. Robustness check: Effects of health shock and other covariates on annual net (log)wage. Injury-purged variables included. Fixed effects estimates.

	(1)	(2)	(3)	(4)
Days-off	-0.1218**	-0.1207**	-0.1258**	-0.1246**
	0.0544	*0.0536*	*0.0574*	*0.0564*
Age	0.9777***	0.9841***	0.9755***	0.9817***
	0.0818	*0.0817*	*0.0819*	*0.0817*
Age sq.	-0.0159***	-0.0160***	-0.0159***	-0.0160***
	0.0013	*0.0013*	*0.0014*	*0.0014*
Goal_purged	0.1150***	0.1150***	0.1149***	0.1148***
	0.0324	*0.0322*	*0.0326*	*0.0323*
Assist_purged	0.0826***	0.0820***	0.0821***	0.0815***
	0.0297	*0.0295*	*0.0297*	*0.0295*
Grade_purged	0.0467**	0.0476**	0.0478**	0.0486**
	0.0202	*0.0201*	*0.0208*	*0.0206*
Popularity	0.0337*	0.0346*	0.0342*	0.0351*
	0.0180	*0.0179*	*0.0181*	*0.0181*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1099	1099	1099	1099
First Stage F-test	*21.46*	*21.98*	*19.77*	*20.34*

***, **, * indicate significance at 1%, 5% and 10%, respectively

[a] Covariates concerning purged Goal, Assist and Grade refers to OLS residuals of the covariate regression on days off, age, age sq. and season fixed effects. [b] The model includes the full set of controls: dummy for the team captain; number of caps in the league during the season; number of senior national team caps during the season; number of national under-21 team caps during the season; dummy for the reoccurrence of the same injury; season fixed effects. [c] Standard Errors in *italics*.

Figure 1. Kernel density estimate of annual net wages

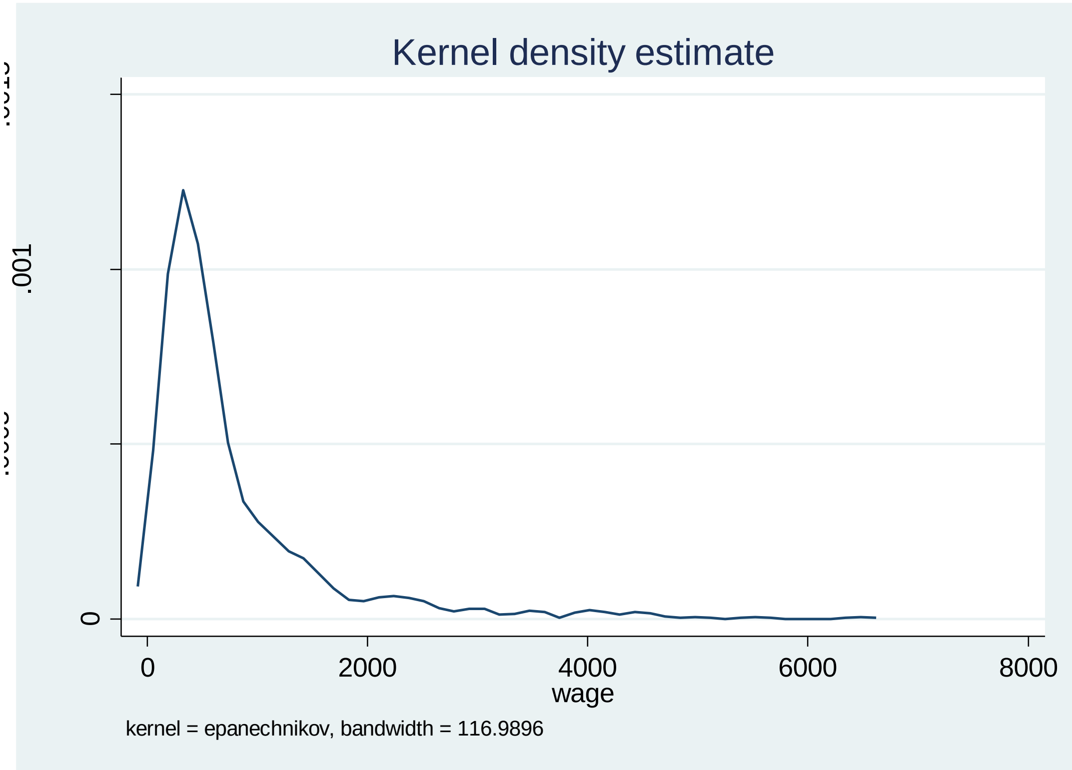
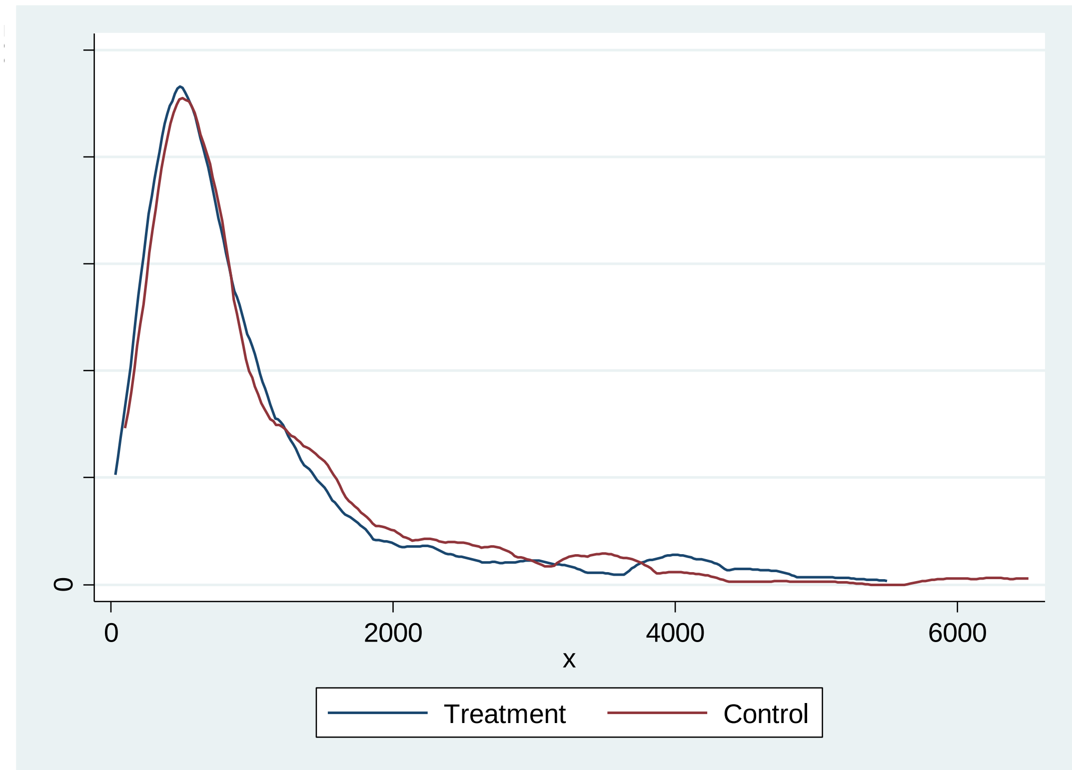


Figure 2. Kernel density estimate of annual net wages: groups comparison



Appendix

Figure A.1 Injuries percentage by month

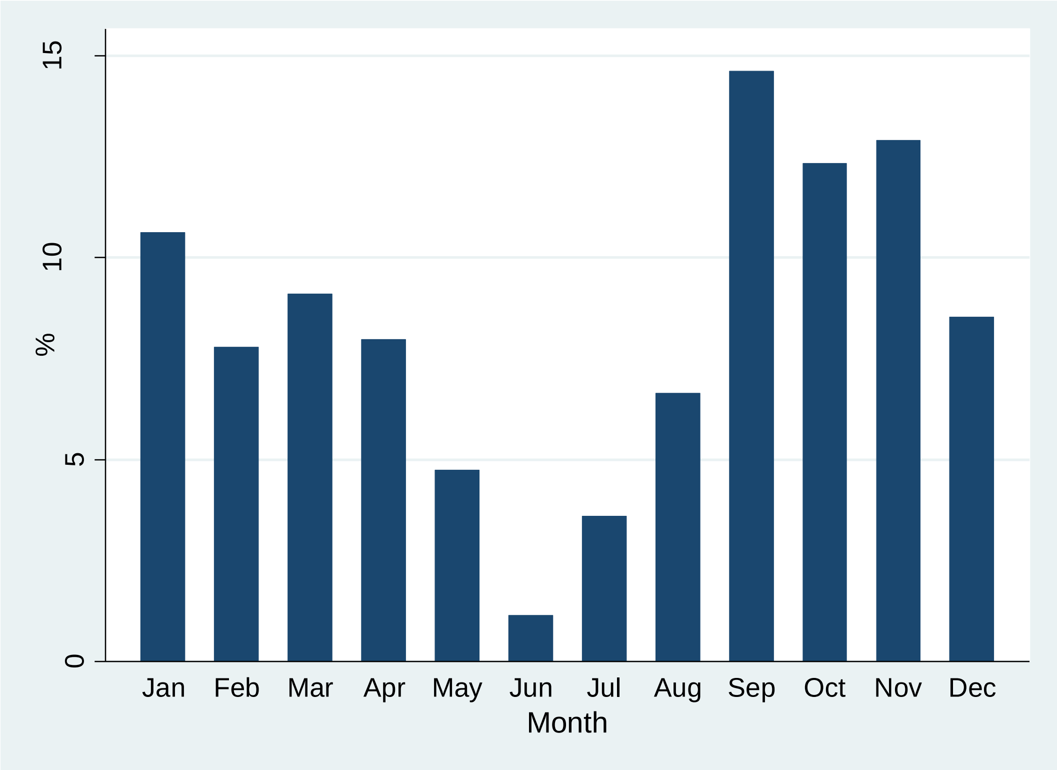


Table A.1 Teams' characteristics. Summary statistics and percentile distribution.

	Wage expense	Net Sales	EBT	Ticketing	Television
Mean	19,656	90,720	-8,630	10,838	53,341
Standard Deviation	13,519	68,682	29,247	10,511	35,240
min	6,9	22,450	-93,767	1,516	7,610
p10	7,85	36,880	-45,919	2,213	25,164
p25	9,1	42,834	-14,040	3880	29,870
p50	13	56,312	-1,745	5,014	34,499
p75	28	116,446	3,636	15,134	70,744
p90	42,95	212,419	12,438	31,017	115,010
p99	47,6	272,404	84,582	38,051	163,478
max	47,6	272,404	84,582	38,051	163,478

[a] Values expressed in thousands of Euros

Table A.2 First stage regressions.

	(1) Days-off	(2) Days-off	(3) Days-off	(4) Days-off
Age	0.7763***	0.7770***	0.7214**	0.7220**
	0.2963	*0.2997*	*0.2947*	*0.2981*
Age sq.	-0.0100**	-0.0100*	-0.0093*	-0.0093*
	0.0051	*0.0051*	*0.0050*	*0.0051*
Under 21	-0.0252	-0.0272	-0.0256	-0.0278
	0.0472	*0.0477*	*0.0467*	*0.0472*
International	0.0096	0.0098	0.0100	0.0101
	0.0157	*0.0159*	*0.0154*	*0.0156*
Captain	-0.1979	-0.1936	-0.2266	-0.2223
	0.3099	*0.3096*	*0.3149*	*0.3145*
Caps	-0.0610***	-0.0613***	-0.0587***	-0.0590***
	0.0083	*0.0084*	*0.0087*	*0.0087*
Goal	0.0113	0.0129	0.0063	0.0079
	0.0851	*0.0851*	*0.0848*	*0.0849*
Assist	0.0385	0.0389	0.0269	0.0271
	0.0629	*0.0632*	*0.0637*	*0.0640*
Grade	0.0749	0.0760	0.0774	0.0787
	0.0793	*0.0795*	*0.0786*	*0.0788*
Popularity	0.0356	0.0355	0.0400	0.0398
	0.0690	*0.0693*	*0.0690*	*0.0694*
Reoccurrence	0.2779	0.2741	0.2859	0.2819
	0.2080	*0.2088*	*0.2111*	*0.2119*
Yellow cards	0.5388***	0.5510***	0.5140***	0.5266***
	0.1595	*0.1585*	*0.1572*	*0.1563*
Team FE	No	Yes	No	Yes
Month FE	No	No	Yes	Yes
N	1099	1099	1099	1099
F-statistic	*20.23*	*20.45*	*18.34*	*18.62*

***, **, * indicate significance at 1%, 5% and 10%, respectively

CHAPTER III

WHO and for how long? An empirical analysis of the consumers' response to red meat warning

Abstract

We explore the effects of the 2015 World Health Organization's warning- about the carcinogenic effect of red meat consumption- on households' behaviour. A striking novelty of my analysis is the examination of both short and long term effects of a health warning and their variations across households differing with respect to average educational level and household health awareness. To identify these effects, I use data collected on a monthly basis that allow to compare variations in a very narrow window across the delivery of the warning and I exploit the strong seasonality patterns in red meat consumption in Italy, mainly associated with culinary traditions in occasion of catholic holidays celebrations. I combine both features in a difference-in-differences framework and I document a general misinterpretation of the warning and a short-term fall in red meat consumption by around 5%. The warning caused a long-lasting and consistent shift in red meat consumption only among households with higher educational levels and health awareness. These findings highlight a brand-new driver of health-education gradient and have relevant implications for the design of public policies delivering health information.

An extract from this chapter is published as: Carrieri, V., & Principe, F. (2018). WHO and for how long? An empirical analysis of the consumers' response to red meat warning (No. 18/08). *HEDG*, c/o Department of Economics, University of York.

1. Introduction

The increasing incidence of illnesses, of which unhealthy diet is one of the key risk factors, represents one of the main health challenges nowadays. According to the WHO (2014), non-communicable diseases related to a poor diet caused 68% of the deaths globally in 2012. Indeed, the poor eating behaviour of the individuals is associated with a vast array of health issues such as obesity, diabetes and cancer, resulting in detrimental effects on individual well-being and leading to poor economic outcomes (Cawley 2015).

In response to this sort of epidemic, the public authorities have increased the volume of information provided about the consequences of unhealthy diet. As documented by the Food and Agriculture Organisation of the United Nations (FAO), there have been increased efforts by international organisations, governments, civil society and the private sector to promote healthy diets in the last twenty years, in both developed and developing countries (Hawkes 2013). The main actions have included media campaigns, nutritional labelling and food safety warnings. However, as with other kinds of information policy, it is quite evident that these initiatives are welfare-improving insofar as they produce a persistent shift in behaviours which is able to generate significant and long-lasting improvements in individual outcomes. As Weiss and Tschirhart (1994) correctly point out, "looking at the effectiveness of public information campaigns directs attention toward the capacity of campaigns to capture the attention of the right audience, to present a clear message, to influence the beliefs or understanding of the audience, and to create the contexts for desired social outcomes". Moreover, insofar as promoting equity is also a twin objective of information activities, it is also important that these activities should be designed in a way of granting accessibility and interpretation also for less-educated groups (Shapiro, 2005).

This study, for the first time in the literature, addresses all these aspects together by investigating the consequences on consumers' behaviour resulting from one important health warning released by the International Agency for Research on Cancer (IARC) of the WHO in October 2015. This was based on a relevant publication appeared on an issue of *The Lancet Oncology* reporting evidence about carcinogenicity of the consumption of red meat and processed meat. In particular, the WHO warning classified some kind of red meat as *Group 2A*, i.e. probably carcinogenic to humans, and processed red meat as *Group 1*, i.e. carcinogenic to humans. The warning is particularly significant as it concerns highly consumed foods which are included in many daily meals around the world. In fact, the news was rapidly circulated by national health authorities, magazines and mass media, and also the demand for information around the topic was rapidly increasing in the period following the warning. Both factors made "red meat" one of the trending topics on the web in October 2015 around the World, (see Section 2 for more details).

We investigate this issue in the geographical context of Italy using data from the Household Budget survey (HBS) which collects expenditures of a large and representative sample of Italian Households. Italy represents an interesting case study and an ideal setting to test these effects for a number of reasons. First, given the long-lasting attention towards food quality associated with the Mediterranean diet. This attention is witnessed, for instance, by a huge amount of related Google searches in the period following the warning; an amount significantly larger than the one observed in almost equally sized countries, such as the UK (see Section 2 for further details). Secondly, available data from Italy includes accurate information on all kinds of expenditure made by a family collected on a diary-form from the 2014 to 2016. Diary based survey is usually taken to be the most reliable way to gather information expenditures and are considered to be of high quality (Browning *et al.* 2003; Browning and Leth-Petersen 2003). Importantly, my data are recorded on a monthly basis. This is a rare feature of expenditure data which are often available only on a quarterly basis. Monthly data allow us to compare households' expenditure variation in a very narrow window across the delivery of the WHO warning and thus to rule out long-term trends in consumption. Lastly, I can exploit a very nice feature of red meat consumption in Italy which is the presence of a strong seasonality in consumption mainly associated with culinary traditions in occasion of catholic holidays celebrations (see Section 2 for more details).

We combine these features in a difference-in-differences framework to analyse both the short and the long-run effect of the WHO warning and their variations across different consumer subgroups, i.e. households differing with respect to average educational level and household health awareness (defined according to the consumption of healthy and unhealthy items, see Section 3). Indeed, when a new piece of health information becomes available, people might respond differently according to their diverse stock of information and ability of processing it as well as to their awareness about the health consequences of certain behaviours (Shapiro, 2005). Moreover, households may need some time to absorb the new pieces of information and to adapt their behaviour and this may lead to very different responses in the short *versus* long run.

This analysis makes a number of contributions to different strands of literature. Firstly, there is a large volume of literature exploring the effects of health authorities' announcements on the households' consumption patterns. Seminal papers (Hamilton 1972; Warner 1989) mostly focused on the smoking hazard campaigns. More recent papers, directly relevant to my study, investigate the effect of food safety advisories on both health and economic outcomes. Smith *et al.* (1988) analyse the impact of media coverage of milk contamination in Hawaii and find that negative news had a greater impact than positive news on consumers' behaviour. Rousu *et al.* (2007) use an experimental design to examine the impact of information about genetically modified food on consumers' willingness to pay. Schlenker and Villas-Boas (2009) found that health warnings about mad cow disease significantly reduced beef sales. Other studies

(Oken *et al.* 2003; Shimshacka *et al.* 2007; Shimshacka and Wardb 2010) document strong evidence of the effects of the 2001 FDA advisory about mercury–related risks in fish consumption. However, the evidence about the effectiveness of public advisories to improve welfare is mixed. On one hand, evidence shows that consumers may under-respond or distrust the advisory (Burger and May, 1996). On the other hand, several studies (e.g. Viscusi 1997; Fox *et al.* 2002) document an alarmist over-reaction to negative information and that consumers tend to place greater weight on more pessimistic sources of risk information. While these studies advance current knowledge on the reactions of consumer to health warnings, they focus essentially on a short-run effect and do not analyse the heterogeneity in the consumer response. This chapter will show that these aspects are, instead, extremely relevant since the response of consumers is actually very heterogeneous across groups both with respect to the magnitude of the effect and to the persistence of the consumption shift.

Secondly, my analysis is linked to the literature exploring the nexus between health policies and preventative behaviour. This literature generally suggests that, consistently with the predictions of rational economic actions (Viscusi *et al.* 1986), the provision of health risk information induces individuals to adopt precautionary behavioural changes. However, with few exceptions (Viscusi *et al.* 1986; Carrieri and Wuebker 2016; Capacci *et al.* 2018), this relies essentially on observational data. However, the identification of a clear causal link between health interventions and risk-taking decisions is difficult in a non-experimental setting, due to unobserved individual characteristics (i.e. risk aversion, inter-temporal preferences) which are likely to affect both the outcome and the individual effort to acquire new information. my quasi-experimental identification strategy allows us to overcome these identification issues and thus to retrieve a causal effect of public policies delivering health information on precautionary behaviours.

Lastly, there is a large body of literature documenting the heterogeneous effects generated by new technology introduction or information availability as a main source of socio-economic status (SES) related health inequalities. In fact, as shown by Contoyannis and Forster (1999), responsiveness to these innovations may vary across socio-economic groups, i.e. a higher take-up rate among the richer or more educated, resulting in a dichotomy between efficiency and equity: average population health and inequalities in health may both increase. As suggested by Deaton (2002) and verified by several empirical papers (Cutler and Lleras-Muney 2006 for a survey, Goesling2007, Conti *et al.* 2010, Clark and Roayer 2013, Lundborg 2013, Brunello *et al.* 2016), education seems to be the key element to disentangle the relationship between socioeconomic status, health outcomes and health innovation uptake. In line with this literature, this study confirms the beneficial effect of education on responsiveness to health warnings. However, it also finds that education along with the health awareness are the main drivers of a stable, more accurate-and not just higher- consumption shift in response to the warning. This may contribute to a better understanding of the role of education and health awareness on SES-related inequalities.

Our findings have relevant implications for the design of public policies delivering health information. In particular, my results suggest that such policies should be designed in a way that expose the individual to a constant flow of information and that particular attention should be paid on the informational message in order to avoid misinterpretation and low compliance especially among less-educated individuals.

The remainder of the chapter is organised as follows. The following section provides more insights into the WHO warning and its media resonance. Section 3 presents the data. In section 4, I discuss my identification strategy. Section 5 presents and discusses the results. Section 6 reports some robustness checks. The last section summarises and concludes.

2. Institutional setting: the WHO warning

In October 2015, the International Agency for Research on Cancer (IARC) of the WHO published an issue of *The Lancet Oncology* reporting evidence about carcinogenicity of the consumption of red meat and processed meat. In particular, red meat was classified as *Group 2A*, i.e. probably carcinogenic to humans, which refers to evidence from epidemiological studies about the association between meat consumption and developing colorectal cancer. On the other hand, processed meet was classified as *Group 1*, i.e. carcinogenic to humans, which refers to sufficient causal evidence linking red meat consumption and cancer in humans. Red meat refers to all mammalian muscle meat, including beef, veal, pork, lamb, mutton, horse, and goat. Processed meat includes meat that has been transformed through salting, curing, fermentation, smoking, or other processes to enhance flavour or improve preservation (e.g., hot dogs, ham, sausages, corned beef and canned meat). According to the IARC, eating 50 grams of processed meat per day increases the risk of colorectal cancer by about 18%, while red meat consumption is associated with an increased risk of developing colorectal, pancreatic, and prostate cancer. These estimates suggests that about 34,000 cancer deaths per year worldwide are attributable to diets high in processed meat (Global Burden of Disease Project 2016); a number that would increase by 50,000 if the relationship with Group 2A red meat was proven to be causal.

Following the evaluation from IARC, the WHO gave health recommendations to prevent the risk of cancer associated with the consumption of meat, inviting individuals to moderate their consumption of meat, particularly processed meat, to reduce the risk of developing cancer. Since the publication of the WHO report in October 2015, the news of the WHO warning had a huge echo across the mass media and was rapidly spread through social networks. To give an idea of this resonance, Figure 1 shows the Google trends for both the search engine hits (as a proxy of the *demand* of information) and the volume of news (the *supply* of information) related to red meat in Italy from 2004 to 2017.

[Figure 1 around here]

As can be seen, both lines representing the relative frequencies, reach their peak in correspondence of October 2015, which is by far the month with the highest volume since 2004 (the first year in which data are available). In Italy, the news had even more echo if compared to countries with a similar population size. For instance, according to the volume data provided by Google AdWords, the term "carne rossa", in Italy, has been searched around 49500 times in October 2015, while its English corresponding "red meat" has been searched only 9600 times in the United Kingdom (a country with an even slightly larger population) in the same period. Interestingly, Figure 1 also shows the presence of other peaks for what concerns the news supply, starting approximately around the middle of 2011. This is attributable to the diffusion of the research outcomes of the first studies exploring the link between the consumption of red meat and some kind cancers, i.e. especially colorectal and prostate cancer (Punnen *et al.* 2011; Takachi *et al.* 2011). However, if in the other cases there was only a consequent negligible increase in the number of search hits by the consumers, the 2015's official warning by the WHO generated by far the highest frequency for both the supply and the consequent demand of information around the health effects of red meat consumption.

3. Data and Variables

Our data come from the Italian Household Budget Survey, which is a cross-sectional survey carried out once a year by the Italian National Institute of Statistics (ISTAT). In agreement with EUROSTAT, the survey is based on the harmonised international classification of expenditure voices (Classification of Individual Consumption by Purpose - COICOP) to ensure international comparability and it is included in the National Statistical Program, i.e. it is used to compute official indicators used by the governments for many purposes such as official relative and absolute poverty thresholds. The survey provides detailed information about the monthly expenditure of the household for goods and services destined for household consumption, alongside a number of demographic and socioeconomic information. Data are collected using a dual system, i.e. a diary followed by a face-to-face interview. In fact, every sampled household receives a diary every month where they are asked to record the daily expenditure sustained by all the household's components, the consumption of goods produced by the household and the place of purchase of goods and services. This information is then validated during the interview in which all other kind of information are also collected. Data are finally made public every year with expenditures listed on a monthly basis. As stressed in the introduction, this is a rare feature of household survey and it will be particularly useful to carefully identify my effects of interest.

In this chapter, I use data from 2014 to 2016. my sample thus consists of about 17,000 observations per wave. Data before 2014 were collected in a different fashion and thus they are not directly comparable to the last two waves. However, information on main aggregates of expenditure are still comparable and I will use them for placebo regressions and to illustrate the validity of the common trend hypothesis (see Section 6 for more details)[25].

Our outcomes, following the IARC's report, refer to the expenditures for the different kind of meats grouped according to their risk classification. Thus, the variable *Group 2A* includes expenditure for beef, pork, lamb and goat; *Group 1* includes cured meat, sausages and canned meat and the variable *Red Meat* includes meats from both groups. Expenditures are expressed in Euros and VAT included.

In the baseline specification, I include the total amount of food expenditure as a control variable. This is in line with the literature about household expenditure (Deaton, 1997) and it is useful to take into account variations over time and between households in the general level of household consumption. As robustness, I also consider a larger set of variables including household demographic and socioeconomic variables: household size, the age range of the household reference person (available in three categories: 18-34, 35-64, 65+), a dummy to indicate whether the household includes migrants and a dummy indicating whether there is at least one graduate in the household. Information about the presence of migrant is useful for taking into account cultural-related food preferences and fasting periods related to religion while the presence of a graduate in the household is useful to take into account both the availability and the ability to process information, which may influence the dietary choices of the entire household. Finally, in order to take into account heterogeneity in regional consumption due to the high local food tradition in Italy, I also include regional fixed effects.

Concerning subgroup analysis, I analyse heterogeneous effects across households with a different level of education and of *health awareness*. Concerning education, I analyse the differential response by households composed by at least one graduate *vs* households with no graduates, while, to assess the role of health awareness, I follow Shimshack *et al.* (2007; 2010) and I consider as *health aware*, the households with a low consumption of unhealthy items such as alcohol and tobacco, i.e. in the bottom two quintiles

[25] Since the 2014, the ISTAT have changed the purpose of the survey, collecting data about expenditures instead of consumption. Moreover, many demographic and socio-economic variables are collected in a very different fashion. As a result, data collected in the waves before 2014 are not directly linkable to the last two waves as explicitly indicated in the data-release documentation.

of the expenditure distribution for these items[26]. A complete description of all these variables along with some descriptive statistics is provided in the next Section.

3. 1 Descriptive statistics

Table 1 shows descriptive statistics of all variables employed in my empirical analysis. Concerning my outcomes, I find that an Italian household spends on average about 78 Euros per month on red meat, while the monthly expenditure for meat included in Group 1 and Group 2A amounts to 34 and 44 Euros, respectively. These expenditures represent 17%, 7%, and 10% of the total expenditure for food, respectively. This confirms the relevance of these items for the Italian household budgets.

[Table 1 around here]

However, average data masks two important features of the expenditure for these items in Italy. These are instead highlighted in Figure 2, which reports the non-parametric distribution of these expenditures. First, I find that the distributions are highly right-skewed. This indicates the presence of very few households consuming high quantities of red meat per month. Second, I find that there is a non-negligible share of households which did not report any expenditure for red meat (about 12% for Group 1 and 18% for Group 2A). Both features are generally common to all households' expenditure data through a log transformation of the dependent variable. An alternative estimation based on Tobit model is presented in Section 6.

[Figure 2 around here]

Concerning the other variables used in my analysis, Table 1 shows that households spend on average 456 Euros per month on food and this represents about the 20% of the total monthly expenditure. In about 20% of the households in my sample there is at least one university graduate and 4% of the households consist of migrants. I also find that households reporting low consumption for unhealthy items such as tobacco and alcohol (i.e. in the bottom two quantiles of the expenditure for these items) spend on average 0.90 Euros per month for these items while their counterparts spend 102 Euros.

Table 2 reports other features of the expenditure on red meat in Italy. First, it highlights the presence of a high regional heterogeneity in the expenditure. Regions in central Italy show higher monthly expenditure in red meat, exceeding by approximately 10 Euros red meat expenditure of Northern and Southern regions. In particular, due to the culinary traditions, Northern regions show higher monthly expenditure in Group 1 meat, while Group 2A meat is more highly consumed in the Southern regions.

[26]Ialso perform some sensitivity across this threshold and results are substantially unchanged (available upon request)

This heterogeneity confirms the need to control for regional fixed effects in my estimates. Interestingly, Table 2 also shows that households reporting low consumption on unhealthy items (alcohol and tobacco) spend less on red meat, i.e. 40 Euros less than their counterparts.

[Table 2 around here]

Figure 3 shows a last interesting feature of the red meat expenditure in Italy, i.e. a strong seasonality. In particular, it emerges that the expenditure is higher in the second and fourth quarter and lower in the first and third quarter of the year. This is likely due to two factors. First, higher expenditure is coincident with the two important Catholic holidays such as Easter and Christmas. In these periods, Italian households cook traditional meals based on red meat, in particular lamb and cured meat, and this explains the peak in consumption during these periods. Second, lower consumption during the summer time (June-September) is likely to be due to the hot temperatures, which make fresh meals based on fruits and vegetables more desirable. An interesting feature of this pattern is that it is highly constant across years. Indeed, Figure 3 shows that monthly and seasonal variations in red meat expenditure are substantially the same in the last two pre-treatment years, i.e. 2013 and 2014. Similar patterns emerge also for other health behaviours such as smoking (Del Bono and Vuri 2017) and it represents a useful source of identification of the causal effects of the WHO warning, as I will show in the next Section.

[Figure 3 around here]

4. Identification strategy

The identification of the effect of the warning on red meat consumption in my setting requires us to address two main challenges. The first challenge is the possible presence of a long-term trend in red meat consumption. Such a trend -especially if negative- would lead to an overestimate of the impact of the warning in a simple before-after framework, as it would confound the effect of the warning with the "natural" trend in red meat consumption. my data released on a monthly basis should allow us to control for this issue since I might compare expenditure variation over a very small window around the time of the release of the warning (i.e. one month before and after the warning) and thus deal with eventual long-term trend effects.[27] However, a potential threat to this strategy might be represented by the existence of a specific shift in red meat consumption after October - other than the one caused by the warning - which

[27] It is important to note that in the short period to which my analysis refers to, the price of red meat was substantially unchanged . According to the official statistics provided by ISTAT, in the 3 months following the WHO's warning the price index comprising red meat reports an average reduction of 0.2% compared to the previous months and an increase of 0.1% compared to the same period of the previous year. These very small variations are then not likely to affect my results.

may bias my effect of interest. To deal with this possibility I exploit the strong seasonality in red meat consumption in Italy mainly associated with culinary traditions in occasion of catholic holidays celebrations, as documented in Section 3.

We take into account these two features in a generalized differences-in-differences framework in which variations in red meat expenditure over a small window around the release of the WHO warning (October 2015) are compared with the variations in the same period of the previous year which actually act as a "control group". A similar identification strategy –but using a larger window around the event- has already been employed in other policy-evaluation frameworks dealing with seasonal effects (i.e. Del Bono and Vuri 2017). An appealing feature of this approach is the possibility of inspect both graphically and with placebo regressions the credibility of the common trend assumption. In my case, this would require a parallel variation in red meat consumption in the months just before and after October over the pre-treatment years. Figure 3 discussed in Section 3 already anticipated that this assumption can be credibly maintained in my setting since seasonal variations in red meat expenditures exhibit a very similar pattern across years. To check this assumption more carefully, in Figure 4Icompare variations in expenditure in November (one month after the warning) with average expenditures in the period January-October in pre-treatment years, i.e. in the 2014 and 2013. Figure 4 shows that these variations are effectively "parallel" , i.e. very similar in 2014 and 2013 and this is found for both kinds of red meat (Group 1 and Group 2A) and also when expenditure for all kinds of red meat is jointly considered. By contrast, significant deviation to this pattern are found in 2015 -the treatment year- as a result of the warning release. This anticipates the existence of a negative and significant consumer response to the WHO warning.

[Figure 4 around here]

In order to further check the validity of this identification assumption, I also perform several placebo regressions using fake warning periods of different length and I implement randomization tests based on simulated placebo warnings for non-parametric inference. Results are reported in Section 6 and support the validity of my identification strategy.

More formally, I estimate the following empirical model:

$$Ln(\mathrm{Meat})_{\mathrm{my}} = \mathrm{T}_{\mathrm{my}}\beta + \theta_{\mathrm{m}} + \gamma_{\mathrm{y}} + \mathrm{u}_{\mathrm{my}} \quad (1)$$

Where the dependent variable is the log of household's expenditure on red meat, Group 1 or Group 2A meat in the month m of the year y, respectively. θ_{m} are month fixed effects capturing monthly variations in expenditure- independently of the year of the interview- while γ_{y} are year fixed effects capturing variations across years in expenditures. T_{my} is my treatment variable and β is the corresponding coefficient measuring the impact of the warning on red meat expenditure. As discussed in the introduction, I also aim to distinguish short *vs* long run effects of the warning. The T_{my} is thus accordingly adapted in different specifications to consider from 1, 2, 5 months and up to one year after October 2015, respectively[28].

In my baseline specification, I also include the total amount of food expenditure as control in equation (1). Alternative specifications includes a larger set of variables, i.e. the household size, the age category of the head of the household, the presence of at least a university graduate in the household and whether the household includes, as discussed in Section 3. I estimate equation (1) using OLS estimator. An alternative estimation using Tobit model is reported in Section 6 and lead to qualitatively similar results.

5. Results

5.1 Short-term effects

Table 3 reports the results of the estimates of the generalized DiD model described in Equation (1) for Red meat, Group 1 and Group 2A meat, respectively. All estimates refer to the short-term effect of the WHO's warning, i.e. one month after the warning took place. In columns 1-3Ireport the estimates of the treatment effect without controls, while in columns 4-6Ireport the estimates of the treatment effect with control variables. For all the outcomes of interest, I report estimates that include standard errors clustered at month level that are robust to correlated monthly shocks in red meat expenditure. However, in Section 6Idemonstrate that my results are robust also to different approaches to statistical inference (block-bootstrap and randomisation tests based on simulated placebo warnings).

[Table 3 around here]

A comparison between columns 1-3 and 4-6 demonstrates that the estimates of the average treatment effect are substantially unchanged when covariates are included. This gives further confidence to the validity of my quasi-experimental design. Table 3 shows that the WHO's warning had a strongly significant impact on consumers' behaviour in the short-term. In fact, in the first month after the treatment, consumers responded to the warning by reducing expenditure on red meat by about 5% (4.9%). Interestingly, the reduction for probably carcinogenic meat (Group 2A) was higher than the one

[28]For this reason I opted to illustrate my estimation equation using a more flexible generalized diff-in-diff formula.

observed for carcinogenic meat (Group 1), amounting on average to 7.4% and 3.2%, respectively. This pattern is likely due to the fact that the news was mainly conveyed through mass-media as a generic "red meat danger" and this induced consumers to reduce especially the consumption of the most known red meats such as beef, pork, lamb and goat. However, as will be shows in the next sub-section, this pattern is highly heterogeneous across households as more educated and health aware families interpreted the warning more correctly, especially in the long-run.

With respect to the control variables, I find that larger households are associated with higher expenditure on red meat, as expected; while, households with at least one university-graduated member spend on average about 10%, 8% and 14% less than less educated households on red, Group 1 and Group 2A meat, respectively. This might be due to a preliminary knowledge around the dangers caused by an excess of red meat consumption which is strengthened by the first research outcomes reporting a correlation between red meat consumption and some kinds of cancer available since 2011 (see the discussion in Section 2). Concerning age, I find that households with an older head of the household spend more on red meat, in particular for what concerns Group 2A meat, for which they spend on average about 12% more than their counterparts. This might be indicative of some cohort effects in red meat consumption. Lastly, as expected, I find that households with migrants are associated with a lower expenditure on red meat and this is likely due to different dietary habits and possibly also related to religious beliefs for some sub-groups of migrants, e.g. Muslims.

5.2 Long-term and heterogeneous effects

In Table 4, I report the estimates of the long-term effect of the warning on households' red meat expenditure. Estimates are based on the same equation described in equation (1) and include the same set of controls but employs a longer post-warning observational period including estimates at two months, five months and one year after the WHO warning, respectively.

[Table 4 around here]

Remarkably, I find that the treatment effect coefficients are negative but never statistically significant at conventional levels in the following months after the release of the warning. This result is consistent across all my outcomes. It is important to observe that testing for the effect up to one year after the warning and accounting for seasonality allows us to reduce any concern about the fact that this result might be influenced by festivity bias and new year's resolutions which might play a role in the adoption of any kind of health behaviour, as already shown by other papers (e.g. Del Bono and Vuri 2017 for smoking; Cherchye *et al.* 2017 for food purchases). Moreover, for sake of brevity, I report in Table 4 only results for two, five months and one year after the warning but additional analyses exclude the presence of any significant treatment effect from two months and up to one year after the release of the warning (results available upon request). Overall, this indicates the presence of a negative causal effect of the warning limited to one month after its release while levels of expenditure in red meat came back to before-warning average levels just two months after its release.

However, these results apply only for the average household and they are indeed extremely heterogeneous across different sub-groups of households, as shown in Tables 5-7. In these Tables, I report both the short and long term estimates of the treatment effects of the WHO's warning for different types of households characterized by a different educational level and *health awareness*. In particular, in the first two rows of each table I report treatment effect estimates for households with at least one university-graduated component *versus* households with no graduate member. As a short-hand I refer to these groups as High-educated *versus* Low-educated households in the Tables. In the third and fourth columns of Tables 5-7Ireport estimates for households with a high *versus* low health awareness (High-aware and Low-aware as a short-hand in the Tables), i.e. households in the top *versus* bottom two quintiles of unhealthy items expenditure such as and alcohol and tobacco. For the sake of brevity, I report in Tables 5-7 only the coefficient measuring the treatment effect. Estimates are based on the same specification discussed so far and includes the same set of controls with the obvious exception of the variables used for sample stratification (i.e. education in the case of comparison between high *vs* low educated households).

[Tables 5, 6, 7 around here]

Concerning education, I find that high-educated households had a stronger and more stable response to the warning. my estimates suggest that these households reduced the expenditure on red meat by about 14% in the first month after the warning, as indicated in Table 5. This reduction is found to be fairly stable over time being equal to about 9% in a span that covers up to one year later the release of the warning. This pattern suggests a sort of permanent shift in red meat consumption for these households. Furthermore, a comparison of results reported in Tables 6 and Table 7 suggests that while in the first month after the warning the reduction was higher for Group 2A meat- in line with the "average household" (as discussed in Section 5.1)- the pattern changes quite substantially when considering long-term effects. Indeed, consumption shifts points towards a higher and stable reduction of carcinogenic Group 1 meat (about 5 Euros in a span that covers up to one year later) and non-significant variations for group 2A meat expenditure. This suggests that high-educated families took time to go deeper into the warning and were able to process the information accurately. This has led to a stable reduction in the more dangerous meat especially in the long run in a way which is consistent with the contents of the WHO warning. This result is not found for low-educated families who instead reduced the generic red meat consumption only in the first month and returned to before-warning expenditure levels just two months after the release of the warning, as shown in the second row of Table 5.

A similar pattern emerges from the comparison of the effects of the warning between households with a different degree of *health awareness*. In order to explore the pure role of health awareness on consumers' response I include household's education level (i.e. the presence of a university graduate in the household) among the set of controls in these estimates. Table 5 shows that Low-aware households had a large reduction of red meat expenditure only in the first month after the warning (about 9 Euros), in line with the average household. On the other hand, High-aware households had an enduring reduction in red meat expenditure, i.e. a constant monthly reduction of about 3 Euros in a span that covers up to one year following the release of the warning. Interestingly, reduction for High-aware households was mostly concentrated on the meat labelled as carcinogen by the WHO (Group 1), as indicated in table 6. Conversely, Low-aware households exhibit a higher reduction among more generic and less dangerous red meat (group 2A) and this is found also in the long run (see Table 7). This indicates, in line with what I observed for education, that health awareness had an impact not only on the magnitude and the stability of the consumption shift but also on the correct adherence to the WHO suggestions.

6. Robustness checks

In this section, I perform a number of checks to test the validity of my identification strategy. As a first check, I focus on the plausibility of the common trend assumption. Figure 3 reported in Section 3 already has shown a seasonality in meat consumption, which was constant across years. A direct consequence of this pattern is that expenditure variations between November and previous months of the year are effectively parallel across time, as shown in Figure 4 of Section 4. Graphical analysis then leads us to be confident about the credibility of common trend assumption in my setting. However, to give even more credence to this assumption, I also replicate the estimates of my Diff-in-Diff regression based on the specification introduced in equation (1) but with "fake warning" periods. In Table 8Ireport placebo DiD estimates assuming a fake warning occurred in October 2014, i.e. exactly one year before the real warning, and using the same post observational period employed for short and long term treatment effects reported in Section 5. We, thus, basically compare the red meat expenditure in the months before the fake warning with periods of up to one year later than the fake warning while accounting for seasonality in red meat expenditure. As expected, the DiD estimates in Table 8 show that treatment variable are never statistically significant alongside all my outcomes and, interestingly, for all post observational period considered (one, two, five months and up to one year later). The coefficients of the control variables are instead very comparable to the ones found in the main regressions reported in Section 5.

We also repeated the same exercise dating the fake warning to two years before, i.e. October 2013, and using a post-observational period of the same length, i.e. up to one year later than the fake warning. Also in this case I do not detect any statistically significant treatment effect. Moreover, I do not detect any significant effect also when performing placebo DiD estimates on the subgroup of households considered for heterogeneous treatment effect estimates as in Section 5.2. (Results are available upon request).

As a second check, I explore the robustness of my results to assumptions about the structure of the error distribution. Indeed, inference in DiD setting might be problematic especially in the presence of a small number of clusters (Bertrand et al., 2004; Donald and Lang, 2007). In my analysis, given the seasonality of the red meat expenditure, the month seems to be the most appropriate level at which to cluster the standard errors. This is the strategy I effectively adopted for the regressions shown in Section 5. Technically, these standard errors are consistent provided that there is a sufficiently large number of clusters. Despite the literature does not offer conclusive evidence around the sufficient number of cluster do draw credible inference, 12 clusters might be effectively "at the boundary". In Table A1 in the Appendix, I show that my results are statistically significant at conventional levels also when based on bootstrapped standard errors clustered at the monthly level (with 200 replications). However, even non-parametric inference based on bootstrapped standard errors might not be entirely reliable with only 12 clusters. Thus, to rule out any possible concern, I follow Bertrand et al. (2004) and I implemented a randomization test based on placebo warnings. Essentially, I randomly select a set of different periods (month x year) for simulating the treatment effect of "fake warnings" and estimate my generalised DiD by using the placebo fake warnings in place of the real one. This process is repeated 2000 times and the estimated coefficients from permutation tests based on Monte Carlo simulations are stored in order to plot the non-parametric distribution of placebo warnings. The main assumption behind this test is that, on average, the fake warning should not generate any effect on the households' red meat expenditure, since the months of treatment effects are randomly chosen.

Figure 5 shows the kernel density distributions of the coefficients generated by the simulation process explained above for my outcomes of interest: red meat, Group 1 and Group 2A meat. As it is possible to observe, the means of the distributions are virtually zero, which implies that estimator of placebo effect is unbiased. More importantly, average treatment effects I estimate for the real WHO's warning (4.9% for red meat, 3.2% for Group 1 and 7.4% for Group 2A as shown in Section 5) fall in the very extreme tails of the distribution of placebo effects. This provides further confidence that the effect I estimated was not observed by chance and therefore reduces any concern about the fact that my results might be incorrect due to invalid assumption on the standard errors distribution.

[Figure 5 around here]

Lastly, I employ a Tobit estimator to check the consistency of my results with respect to the excess of zeros problem, a common alternative in the empirical literature on the analysis of expenditure data (e.g. Donkers et al. 2017, for charity expenditure; Tansel and Bircan 2006, for education expenditure; and Cai 1998, for food expenditure).Estimates are reported in Table 9 and result to be qualitatively unchanged when compared to my main estimates.

[Table 9 around here]

The coefficients of the Tobit model encompass both changes in the probability of having positive expenditure on red meat and changes in red meat expenditure for those with a positive expenditure in red meat. Thus , I apply the decomposition method suggested by McDonald and Moffitt (1980) which allows us to assess the relative weight of these two effects. I find that 73% of the total change in expenditure on red meat was generated by marginal changes in the value of positive expenditures, whereas 27% was generated by changes in the probability of spending anything at all for red meat. This is consistent with the contents of the WHO warning which was that of reducing rather than eliminating red meat consumption. These effects are 65% and 35% for Group 1 and 60% and 40% for Group 2A, respectively[29]. Interestingly, this indicates that warning seems to have increased more the proportion of consumers who do not consume any quantity of less dangerous red meat (Group 2A) than carcinogenic meat (Group 1). This confirms the common misinterpretation of the generic "red meat danger" previously discussed.

7. Conclusions

In this chapter, I have investigated the consequences on consumers' behaviour of World Health Organization's warning concerning the carcinogen effect of the red meat consumption released in October 2015. I investigated this topic in Italy due to its the long-lasting attention towards food quality and Mediterranean diet but also for the availability of good data collecting expenditures for a large and representative sample of Italian Households on monthly basis and with rich information on household characteristics. Monthly data allow us to compare households' expenditure variation in a very narrow window across the delivery of the WHO warning and thus to rule out long-term trends in consumption.

[29] Mc Donald and Moffit (1980) decompose the total effect of a determinant $\boldsymbol{X_i}$ in a tobit model as: $\boldsymbol{\delta Ey/\delta X_i = F(z)(\delta Ey^*/\delta X_i) + Ey^*(\delta F(z)/\delta X_i)}$, where $\boldsymbol{F(z)}$ is the share of observations with non-zero expenditures, $\boldsymbol{\delta Ey^*/\delta X_i}$ is the impact of the determinant on the expenditure above zero, $\boldsymbol{Ey^*}$is the average positive expenditure and $\boldsymbol{\delta F(z)/\delta X_i}$ is the impact of the determinant on the probability of any expenditure. In the case of red meat, the decomposition is $\boldsymbol{5.24 = (0.93 * 3.9) + (80.9 * 0.02)}$. Thus, the first term (0.93*3.9)/5.24) accounts for 73% of the final coefficient, while the remaining 27% is explained by the second term (80.9*0.02/5.24). Same procedure was applied for the other outcomes

Moreover, I exploit a very nice feature of red meat consumption in Italy which is the presence of a strong seasonality in consumption mainly associated with culinary traditions in occasion of catholic holidays celebrations. I combine both features in a difference-in-differences framework that allows to retrieve the causal effect of the warning on red meat expenditure under the hypothesis of common trend in expenditure over the same period of the year, which seems to be largely supported in my case. Additionally, the availability of data up until one year after the warning and detailed information around household's characteristics and expenditures allows us to analyse both the short and the long-run effect of the WHO warning and their variation across different consumers' subgroups.

Using the Household Budget Survey (HBS), my analysis leads to three important findings. Firstly, I find that WHO's warning had a strongly significant impact on consumer's behaviour but only in the very short-term. In fact, I find that in the first month after the warning, consumers reduced their expenditure on red meat by around 5%, 3.2% and 7.4% of the average monthly expenditure in generic red meat, in carcinogenic meat (*Group 1)* and in probably carcinogenic meat (*Group 2A)*, respectively. However, expenditures on red meat returned to pre-warning levels just two months after its release. Secondly, I find that only some subgroups, such as the more educated and health aware changed their eating behaviours in the long run, i.e. over a one year post-warning observational period. On the contrary, poor-educated and less health-aware households reduced the consumption in a less significant manner and only in the very short-term. Thirdly, I also find that these groups differ significantly with respect to the correct interpretation of the warning. More educated and health-aware households reduced especially the consumption of carcinogenic meat (Group 1) while their counterparts reduced mostly the consumption of relatively less dangerous meat (Group 2A). This pattern is likely due to the fact that the news was mainly conveyed through mass-media as a generic "red meat danger" and this induced these kinds of consumers to reduce particularly their consumption of the most common but relatively less dangerous red meats such as beef, pork, lamb and goat.

These results contribute to three different strands of the literature and offer potentially very relevant implications around the design of food warning policies. Firstly, I contribute to a large volume of literature exploring the effects of health authorities' announcement on households' consumption patterns. I add to this literature by demonstrating that the effect of an announcement might be very different in the short *versus* long run and highly heterogeneous across subgroups of consumers. Secondly, I report causal evidence on the effect of health information on risk-taking decisions in a quasi-experimental setting. This topic has been addressed by other papers based on observational data which however cannot tackle the endogeneity between the consumers' efforts to acquire information and their behaviours. Thirdly, I contribute to the literature exploring the distributive consequences of new technology introduction or information availability. In line with this literature, I confirm that education plays a significant role in the responsiveness to health warnings. This indicates that both the stock of information and the ability to process it are factors which contribute to shaping health inequalities in a substantial way. However, I also find a new-brand driver of the health-education gradient. Indeed, education along with the health awareness are found to be the drivers of a stable, more accurate -and not just higher- consumption shift in response to the informational shock. This offers a perhaps more pessimistic view on the possibility of contrasting health inequalities through educational campaigns especially when the aim is to change behaviours in a permanent way.

In terms of policy, my paper has a number of implications for the successful design of health advices. Firstly, the fact that the consumers - on average- only responded in the immediate short-term suggests that successful health information policies should be designed in a way that expose the individual to a constant flow of information. This is likely to be more beneficial compared to the "one-shot" warnings. A real-world example of these policies is represented by the anti-smoking advice campaigns, which are delivered quite constantly, and through different formats. Secondly, the misinterpretation of the warnings by some subgroups suggests the importance of re-designing the health warnings in a simpler and more accessible way. In this respect, a higher involvement of health care providers in dissemination of health warnings might be beneficial. Other than being a likely cost-effective strategy to increase general adherence to the health warnings, as argued by Chaloupka et al., (2011), this might also have the spillover effect of reducing the information gap between more and less educated. Finally, my findings confirm the strategic role played by education for health. Other than to reduce the well-known health gap, my finding indirectly suggests also that education is able to increase the health returns on investments in health campaigns and health educational activities since the latter are misinterpreted by low educated individuals and produce only short-term effects among them. In a general equilibrium perspective, higher investments in education are then likely to bring both equity and efficiency gains to the health production process.

References

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much shouldItrust differences-in-differences estimates?. *The Quarterly Journal of Economics*, *119*(1), 249-275.

Browning, M., Crossley, T. F., & Weber, G. (2003). Asking consumption questions in general purpose surveys. *The Economic Journal*, *113*(491), F540-F567.

Browning, M., & Leth‐Petersen, S. (2003). Imputing consumption from income and wealth information. *The Economic Journal*, *113*(488), F282-F301.

Brunello, G., Fort, M., Schneeweis, N., & Winter‐Ebmer, R. (2016). The causal effect of education on health: What is the role of health behaviors?. *Health economics*, *25*(3), 314-336.

Burger, J.,& May, H. (1996). Fishing in a polluted estuary: fishing behavior, fish consumption, and potential risk. *Risk Analysis*, *16*(4), 459-471.

Cai, L. A. (1998). Analyzing household food expenditure patterns on trips and vacations: a Tobit model. *Journal of Hospitality & Tourism Research*, *22*(4), 338-358.

Capacci, S., Mazzocchi, M., & Shankar, B. (2018). Breaking Habits: The Effect of the French Vending Machine Ban on School Snacking and Sugar Intakes. *Journal of Policy Analysis and Management*, *37*(1), 88-111.

Carrieri, V., & Wuebker, A. (2016). Quasi‐Experimental Evidence on the Effects of Health Information on Preventive Behaviour in Europe. *Oxford Bulletin of Economics and Statistics*, *78*(6), 765-791.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of health economics*, *43*, 244-268.

Chaloupka, F.J., Powell, L.M, & Chirqui, J.F. (2011). Sugar-sweetened beverages and obesity: The potential impact of public policies. *Journal of Policy Analysis and Management*, 30 (3), 645-655

Cherchye, L., De Rock, B., Griffith, R., O'Connell, M., Smith, K., & Vermeulen, F. (2017). A new year, a new you? Heterogeneity and self-control in food purchases. *IZA Discussion Papers*, n. 11205.

Clark, D., & Roayer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *The American Economic Review*, *103*(6), 2087-2120.

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *The American Economic review*, *100*(2), 234-238.

Contoyannis, P., & Forster, M. (1999). The distribution of health and income: a theoretical framework. *Journal of Health Economics*, *18*(5), 605-622.

Cutler, D. M., & Lleras-Muney, A. (2006). Education and health: evaluating theories and evidence (No. w12352). National bureau of economic research.

Deaton, A. (1997). The analysis of household surveys: a microeconometric approach to development policy. World Bank Publications.

Deaton, A. (2002). Policy implications of the gradient of health and wealth. *Health Affairs*, *21*(2), 13-30.

Del Bono, E., & Vuri, D. (2017). Smoking behaviour and individual well-being: a fresh look at the effects of the 2005 public smoking ban in Italy. *Oxford Economic Papers*.

Donkers, B., van Diepen, M., & Franses, P. H. (2017). Do charities get more when they ask more often? Evidence from a unique field experiment. *Journal of Behavioral and Experimental Economics*, *66*, 58-65.

Fox, J. A., Hayes, D. J., & Shogren, J. F. (2002). Consumer preferences for food irradiation: How favorable and unfavorable descriptions affect preferences for irradiated pork in experimental auctions. *Journal of Risk and Uncertainty*, *24*(1), 75-95.

Goesling, B. (2007). The rising significance of education for health?. *Social Forces*, *85*(4), 1621-1644.

Hamilton, J. L. (1972). The demand for cigarettes: advertising, the health scare, and the cigarette advertising ban. *The Review of Economics and Statistics*, 401-411.

Hawkes C. 2013. Promoting healthy diets through nutrition education and changes in the food environment: an international review of actions and their effectiveness. Rome: Nutrition Education and Consumer Awareness Group, Food and Agriculture Organization of the United Nations.

Lundborg, P. (2013). The health returns to schooling—what canIlearn from twins?. *Journal of Population Economics*, *26*(2), 673-701.

Oken, E., Kleinman, K. P., Berland, W. E., Simon, S. R., Rich-Edwards, J. W., & Gillman, M. W. (2003). Decline in fish consumption among pregnant women after a national mercury advisory. *Obstetrics & Gynecology*, *102*(2), 346-351.

Punnen, S., Hardin, J., Cheng, I., Klein, E. A., & Witte, J. S. (2011). Impact of meat consumption, preparation, and mutagens on aggressive prostate cancer. *PLoS One*, *6*(11), e27711.

Rousu, M., Huffman, W. E., Shogren, J. F., & Tegene, A. (2007). Effects and value of verifiable information in a controversial market: evidence from lab auctions of genetically modified food. *Economic Inquiry*, *45*(3), 409-432.

Shapiro, M.D., (2005). Equity and information: Information regulation, environmental justice, and risks from toxic chemicals. *Journal of Policy Analysis and Management*, 24(2), 373-398.

Schlenker, W., & Villas-Boas, S. B. (2009). Consumer and market responses to mad cow disease. *American Journal of Agricultural Economics*, *91*(4), 1140-1152.

Shimshack, J. P., & Ward, M. B. (2010). Mercury advisories and household health trade-offs. *Journal of Health Economics*, *29*(5), 674-685.

Shimshack, J. P., Ward, M. B., & Beatty, T. K. (2007). Mercury advisories: information, education, and fish consumption. *Journal of Environmental Economics and Management*, *53*(2), 158-179.

Smith, M. E., Van Ravenswaay, E. O., & Thompson, S. R. (1988). Sales loss determination in food contamination incidents: an application to milk bans in Hawaii. *American Journal of Agricultural Economics*, *70*(3), 513-520.

Takachi, R., Tsubono, Y., Baba, K., Inoue, M., Sasazuki, S., Iwasaki, M., & Tsugane, S. (2011). Red meat intake may increase the risk of colon cancer in Japanese, a population with relatively low red meat consumption. *Asia Pacific journal of clinical nutrition*, *20*(4), 603-612.

Tansel, A., & Bircan, F. (2006). Demand for education in Turkey: A tobit analysis of private tutoring expenditures. *Economics of education review*, *25*(3), 303-313.

Viscusi, W. K. (1997). Alarmist decisions with divergent risk information. *The Economic Journal*, *107*(445), 1657-1670.

Viscusi, W. K., Magat, W. A., & Huber, J. (1986). Informational regulation of consumer health risks: an empirical evaluation of hazard warnings. *The Rand Journal of Economics*, 351-365.

Warner, K. E. (1989). Effects of the antismoking campaign: an update. *American Journal of Public Health*, *79*(2), 144-151.

Weiss, J.A., Tschirhart, M. (1994). Public information campaigns as policy instruments. *Journal of Policy Analysis and Management*, 13(1), 82-119.

WHO. (2014). Global Status Report on Non-communicable Diseases. Geneva: *World Health Organization*.

Tables and Figures

Table 1. Summary statistics

Variable	*Description*	*Mean*	*Std. Dev.*
Dependent variables			
Red Meat	Monthly expenditure on Red meat	78.29	69.98
Group 1	Monthly expenditure on Group 1 meat	34.00	32.73
Group 2A	Monthly expenditure on Group 2A meat	44.29	49.48
Controls/Subgroups			
Food expenditure	Monthly expenditure on food	456.54	304.48
HH size	Household size	2.35	1.22
High-Educated	At least one graduate in the household (share)	0.21	0.40
Age	Age category of the household's respondent	18-34 (7%) 35-64 (55%) >64 (38%)	
Migrant	At least one migrant in the household (share)	0.043	0.20
High-aware	Average expenditure of households in the bottom two quintiles of Tobacco and Alcohol expenditure	175.48	82.17

All expenditure values are expressed in Euros

Table 2. Meat expenditure by subgroup: mean values

	Red Meat	Group1	Group 2A
All	78.29	34.00	44.29
North	76.68	35.46	41.22
Centre	85.42	36.11	49.31
South	75.74	30.73	45.00
High Education	82.79	36.37	46.42
Low Education	77.12	33.38	43.73
High-aware	99.16	42.68	56.47
Low-aware	58.11	25.48	32.63

All values expressed in Euros

Table 3. DiD estimates of the effect of the warning on meat expenditure: short-term effects

	(1) Red	(2) Group 1	(3) Group 2A	(4) Red	(5) Group 1	(6) Group 2A
Treatment	-0.0491***	-0.0320***	-0.0742***	-0.0478***	-0.0313***	-0.0722***
	0.0056	*0.0083*	*0.0091*	*0.0055*	*0.0080*	*0.0091*
Food Exp.	0.0015***	0.0013***	0.0017***	0.0014***	0.0012***	0.0016***
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
H Size				0.0527***	0.0530***	0.0502***
				0.0022	*0.0032*	*0.0032*
High-Educ.				-0.1049***	-0.0833***	-0.1388***
				0.0073	*0.0088*	*0.0092*
HH Age 35-65				0.0241**	-0.0025	0.0525**
				0.0090	*0.0109*	*0.0187*
HH Age>65				0.0446***	-0.0424**	0.1239***
				0.0127	*0.0147*	*0.0198*
Migrant				-0.1503***	-0.2204***	-0.1054***
				0.0130	*0.0188*	*0.0180*
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
N	30852	30852	30852	30852	30852	30852

OLS estimates of Equation (1). Clustered standard errors at month level in *italics*. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table 4. DiD estimates of the effect of the warning on meat expenditure: long-term effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Red			Group 1			Group 2A	
	2-month	5-month	1-year	2-month	5-month	1-year	2-month	5-month	1-year
Treatment	-0.0136	-0.0137	-0.0141	0.0038	0.0039	0.0038	-0.0322	-0.0325	-0.0330
	0.0230	*0.0231*	*0.0226*	*0.0243*	*0.0244*	*0.0238*	*0.0278*	*0.0277*	*0.0275*
Food Exp.	0.0014***	0.0014***	0.0014***	0.0012***	0.0012***	0.0012***	0.0016***	0.0016***	0.0016***
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
H size	0.0528***	0.0513***	0.0515***	0.0522***	0.0503***	0.0503***	0.0509***	0.0498***	0.0498***
	0.0021	*0.0026*	*0.0022*	*0.0032*	*0.0033*	*0.0036*	*0.0032*	*0.0033*	*0.0028*
High-Educ.	-0.1081***	-0.1095***	-0.1093***	-0.0871***	-0.0881***	-0.0940***	-0.1411***	-0.1433***	-0.1385***
	0.0074	*0.0072*	*0.0067*	*0.0079*	*0.0078*	*0.0081*	*0.0093*	*0.0094*	*0.0094*
HH Age 34-64	0.0310**	0.0309**	0.0342**	0.0017	0.0034	0.0036	0.0602**	0.0601**	0.0685***
	0.0133	*0.0123*	*0.0118*	*0.0118*	*0.0105*	*0.0087*	*0.0215*	*0.0198*	*0.0172*
HH Age>65	0.0518***	0.0519***	0.0542***	-0.0399**	-0.0397**	-0.0396***	0.1336***	0.1345***	0.1430***
	0.0166	*0.0154*	*0.0130*	*0.0150*	*0.0139*	*0.0109*	*0.0239*	*0.0219*	*0.0187*
Migrant	-0.1450***	-0.1387***	-0.1412***	-0.2145***	-0.2051***	-0.2039***	-0.1019***	-0.0968***	-0.1022***
	0.0147	*0.0131*	*0.0101*	*0.0210*	*0.0209*	*0.0158*	*0.0173*	*0.0145*	*0.0158*
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	32782	35814	45837	32782	35814	45837	32782	35814	45837

OLS estimates of Equation (1). Clustered standard errors at month level in italics. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table 5. DiD estimates of the effect warning on red meat expenditure: heterogeneous effects

	(1) 1-month	(2) 2-month	(3) 3-month	(4) 4-month	(5) 5-month	(6) 1-year
High-Educ.	-0.1430***	-0.0882**	-0.0882**	-0.0871**	-0.0879**	-0.0890**
	0.0114	*0.0367*	*0.0365*	*0.0365*	*0.0364*	*0.0362*
Low-Educ.	-0.0254***	0.0048	0.0048	0.0048	0.0047	0.0048
	0.0069	*0.0212*	*0.0212*	*0.0212*	*0.0213*	*0.0210*
Low-aware	-0.0836***	-0.0300	-0.0301	-0.0301	-0.0305	-0.0316
	0.0117	*0.0336*	*0.0335*	*0.0335*	*0.0336*	*0.0332*
High-aware	-0.0281***	-0.0352***	-0.0349***	-0.0351***	-0.0351***	-0.0349***
	0.0085	*0.0094*	*0.0095*	*0.0096*	*0.0095*	*0.0095*

OLS estimates coefficients of treatment effect of Equation (1). Full set of controls included. Clustered standard errors at month level in *italics*. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table 6. DiD estimates of the effects warning on Group 1 meat expenditure: heterogeneous effects

	(1) 1-month	(2) 2-month	(3) 3-month	(4) 4-month	(5) 5-month	(6) 1-year
High-Educ.	-0.0780***	-0.0678***	-0.0679***	-0.0669***	-0.0676***	-0.0682***
	0.0154	*0.0163*	*0.0162*	*0.0161*	*0.0159*	*0.0153*
Low-Educ.	-0.0208**	0.0227	0.0228	0.0227	0.0227	0.0231
	0.0088	*0.0300*	*0.0301*	*0.0302*	*0.0302*	*0.0297*
Low-aware	-0.0660***	-0.0119	-0.0119	-0.0122	-0.0126	-0.0124
	0.0143	*0.0355*	*0.0355*	*0.0354*	*0.0353*	*0.0349*
High-aware	-0.0206**	-0.0233**	-0.0227**	-0.0228**	-0.0228**	-0.0226**
	0.0089	*0.0087*	*0.0088*	*0.0089*	*0.0089*	*0.0092*

OLS estimates coefficients of treatment effect of Equation (1). Full set of controls included. Clustered standard errors at month level in *italics*. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table 7. DiD estimates of the effect of the warning on Group 2A meat expenditure: heterogeneous effects

	(1) 1-month	(2) 2-month	(3) 3-month	(4) 4-month	(5) 5-month	(6) 1-year
High-Educ.	-0.1980***	-0.1059	-0.1059	-0.1049	-0.1062	-0.1069
	0.0139	*0.0609*	*0.0606*	*0.0606*	*0.0607*	*0.0604*
Low-Educ.	-0.0430***	-0.0153	-0.0155	-0.0153	-0.0155	-0.0157
	0.0110	*0.0217*	*0.0217*	*0.0216*	*0.0217*	*0.0217*
Low-aware	-0.1061***	-0.0467	-0.0469	-0.0466	-0.0472	-0.0492
	0.0143	*0.0375*	*0.0373*	*0.0374*	*0.0376*	*0.0371*
High-aware	-0.0554***	-0.0573***	-0.0573***	-0.0577***	-0.0578***	-0.0571***
	0.0136	*0.0134*	*0.0135*	*0.0137*	*0.0137*	*0.0138*

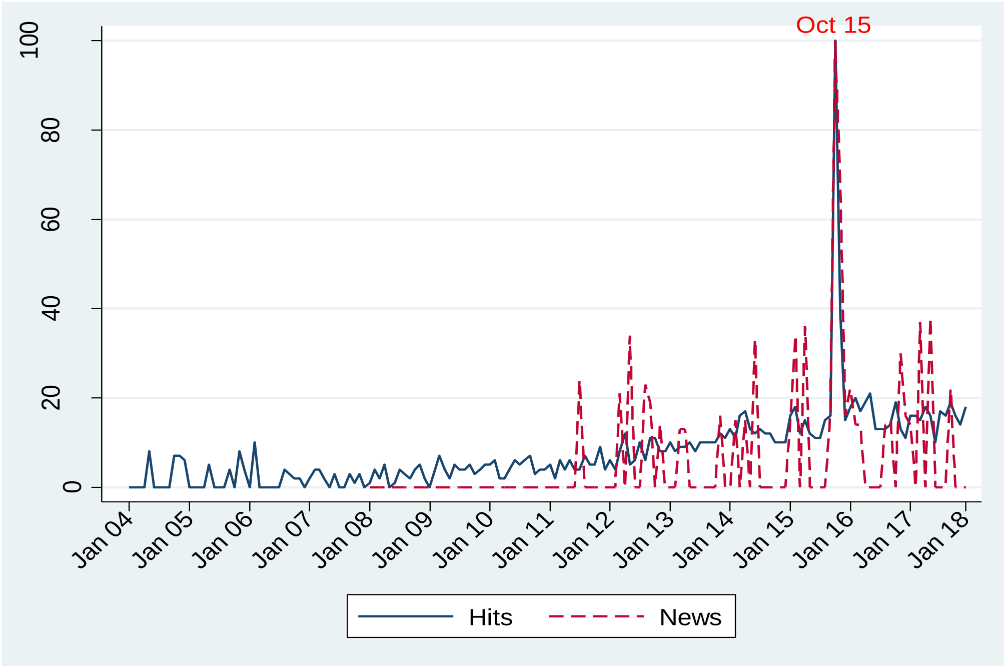
OLS estimates coefficients of treatment effect of Equation (1). Full set of controls included. Clustered standard errors at month level in *italics*. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Table 8. Robustness checks. Placebo tests for fake warning periods

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Red meat				Group 1				Group 2A			
	1-month	2-month	5-month	1-year	1-month	2-month	5-mont	1-year	1-month	2-month	5-mont	1-year
Treatment	0.0043	-0.0094	-0.0098	-0.0101	-0.0042	-0.0267	-0.0269	-0.0267	0.0038	-0.0008	-0.0014	-0.0018
	_0.0044_	_0.0114_	_0.0112_	_0.0116_	_0.0087_	_0.0194_	_0.0191_	_0.0186_	_0.0075_	_0.0083_	_0.0081_	_0.0085_
Food Exp.	0.0014***	0.0014***	0.0014***	0.0014***	0.0013***	0.0013***	0.0013***	0.0013***	0.0017***	0.0017***	0.0016***	0.0016***
	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_	_0.0000_
H Size	0.0419***	0.0423***	0.0433***	0.0445***	0.0491***	0.0490***	0.0493***	0.0508***	0.0351***	0.0359***	0.0375***	0.0388***
	_0.0032_	_0.0030_	_0.0030_	_0.0024_	_0.0031_	_0.0030_	_0.0028_	_0.0029_	_0.0055_	_0.0053_	_0.0048_	_0.0038_
High Educ.	-0.0428***	-0.0450***	-0.0503***	-0.0572***	-0.0166*	-0.0154*	-0.0191**	-0.0312***	-0.0675***	-0.0723***	-0.0809***	-0.0854***
	_0.0072_	_0.0067_	_0.0054_	_0.0049_	_0.0077_	_0.0074_	_0.0065_	_0.0065_	_0.0108_	_0.0103_	_0.0085_	_0.0080_
HH Age 35-64	0.0155*	0.0193*	0.0187**	0.0221***	0.0008	0.0027	0.0015	0.0054	0.0263*	0.0312**	0.0315**	0.0352**
	_0.0077_	_0.0090_	_0.0080_	_0.0066_	_0.0142_	_0.0146_	_0.0129_	_0.0093_	_0.0138_	_0.0140_	_0.0121_	_0.0144_
HH Age>65	0.0421***	0.0466***	0.0445***	0.0467***	-0.0420***	-0.0379**	-0.0419**	-0.0366***	0.0991***	0.1041***	0.1026***	0.1049***
	_0.0084_	_0.0096_	_0.0087_	_0.0073_	_0.0134_	_0.0152_	_0.0141_	_0.0113_	_0.0162_	_0.0158_	_0.0139_	_0.0145_
HH Migrant	-0.1552***	-0.1507***	-0.1607***	-0.1484***	-0.2963***	-0.2794***	-0.2858***	-0.2735***	-0.0916***	-0.0925***	-0.1041***	-0.0834***
	_0.0124_	_0.0122_	_0.0114_	_0.0120_	_0.0211_	_0.0264_	_0.0267_	_0.0206_	_0.0154_	_0.0140_	_0.0139_	_0.0164_
_N_	36095	37421	41414	50351	36095	37421	41414	50351	36095	37421	41414	50351

OLS estimates of Equation (1) for fake warning (October 2014). Full set of controls included. Clustered standard errors at month level in _italics_. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

Figure 1. Google trends for "carne rossa" (red meat) in Italy, 2004-2018



Own elaboration on Google trends data. Google trends data for News are only available from 2008.

Figure 2. Kernel density estimate of monthly expenditure on red meat

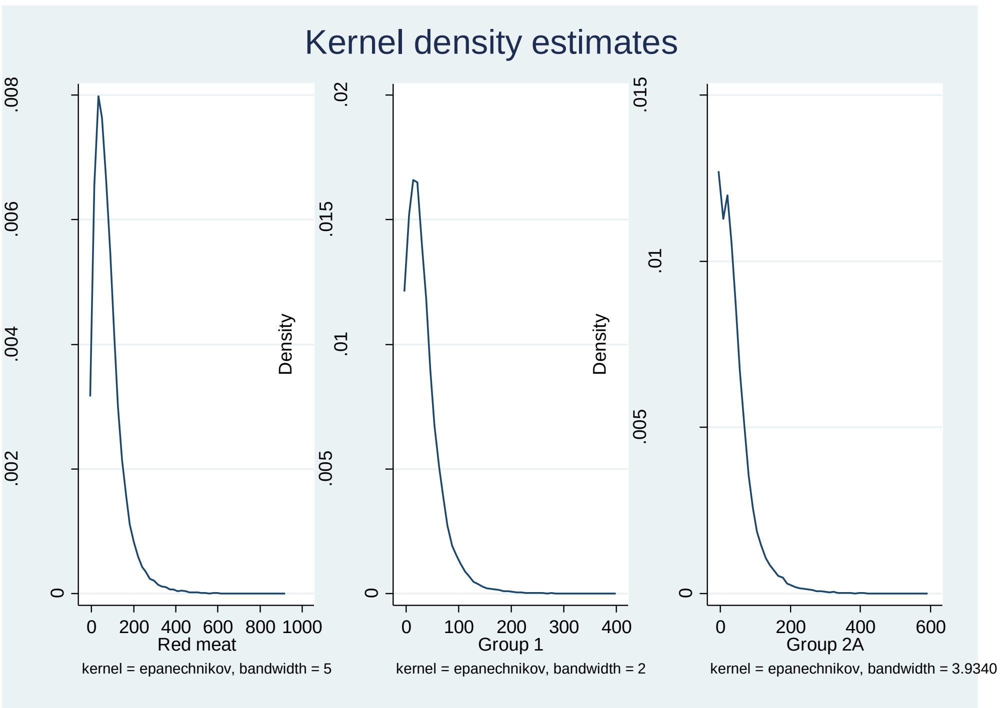


Figure 3. Share of red meat expenditure by month and quarter, 2013 and 2014

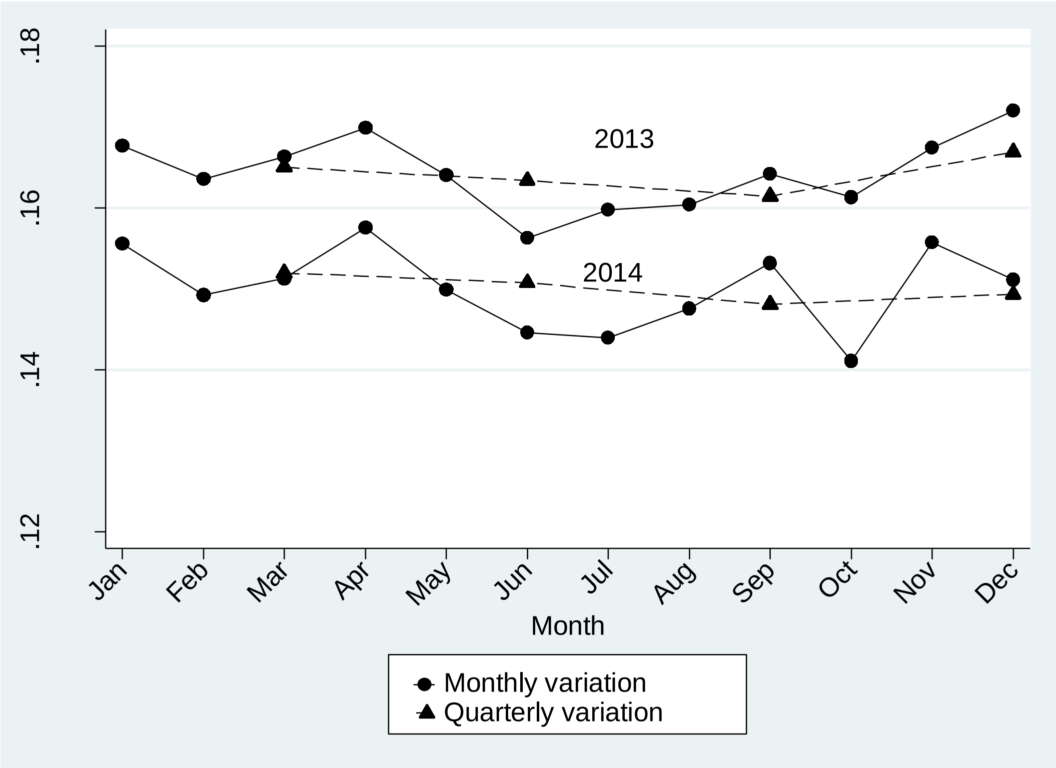


Figure 4. Common trends in red meat expenditure

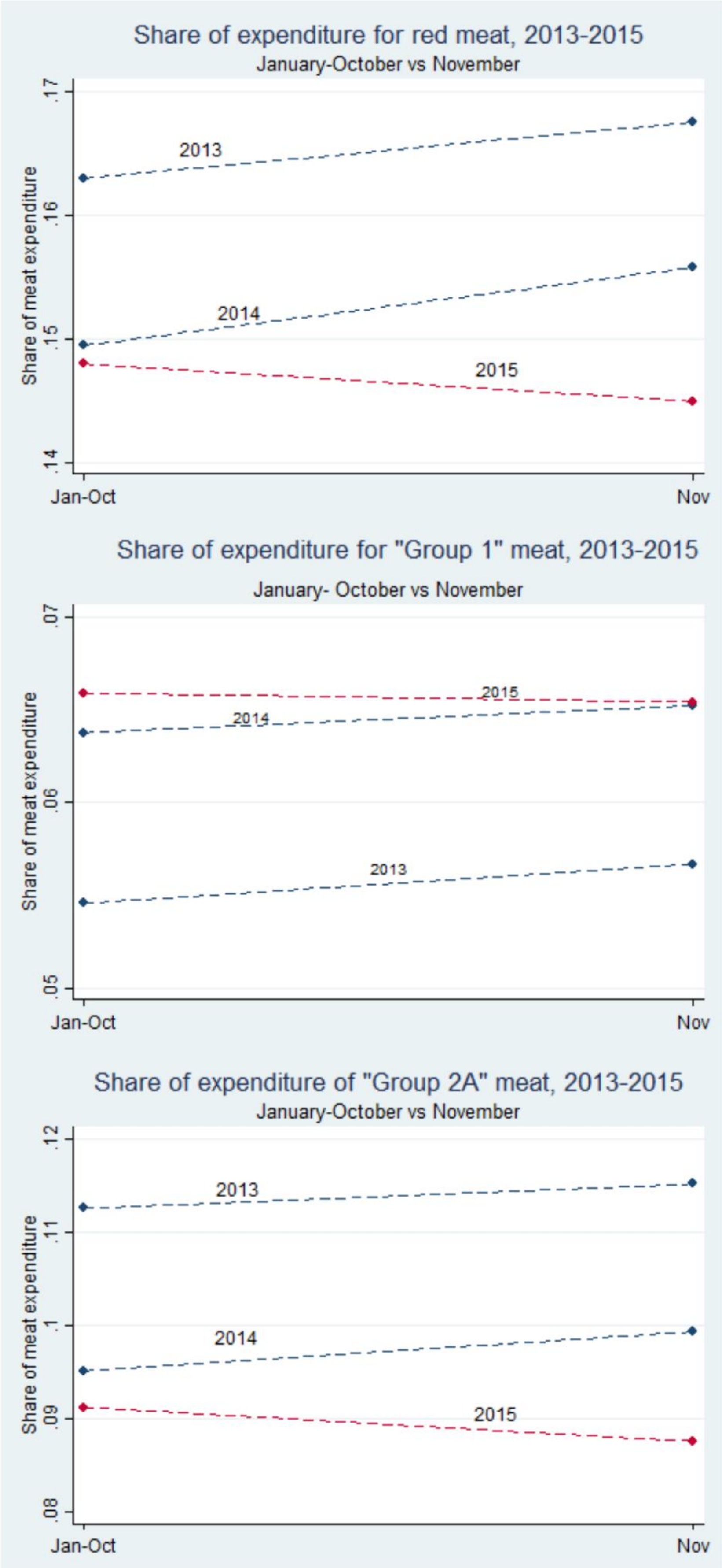
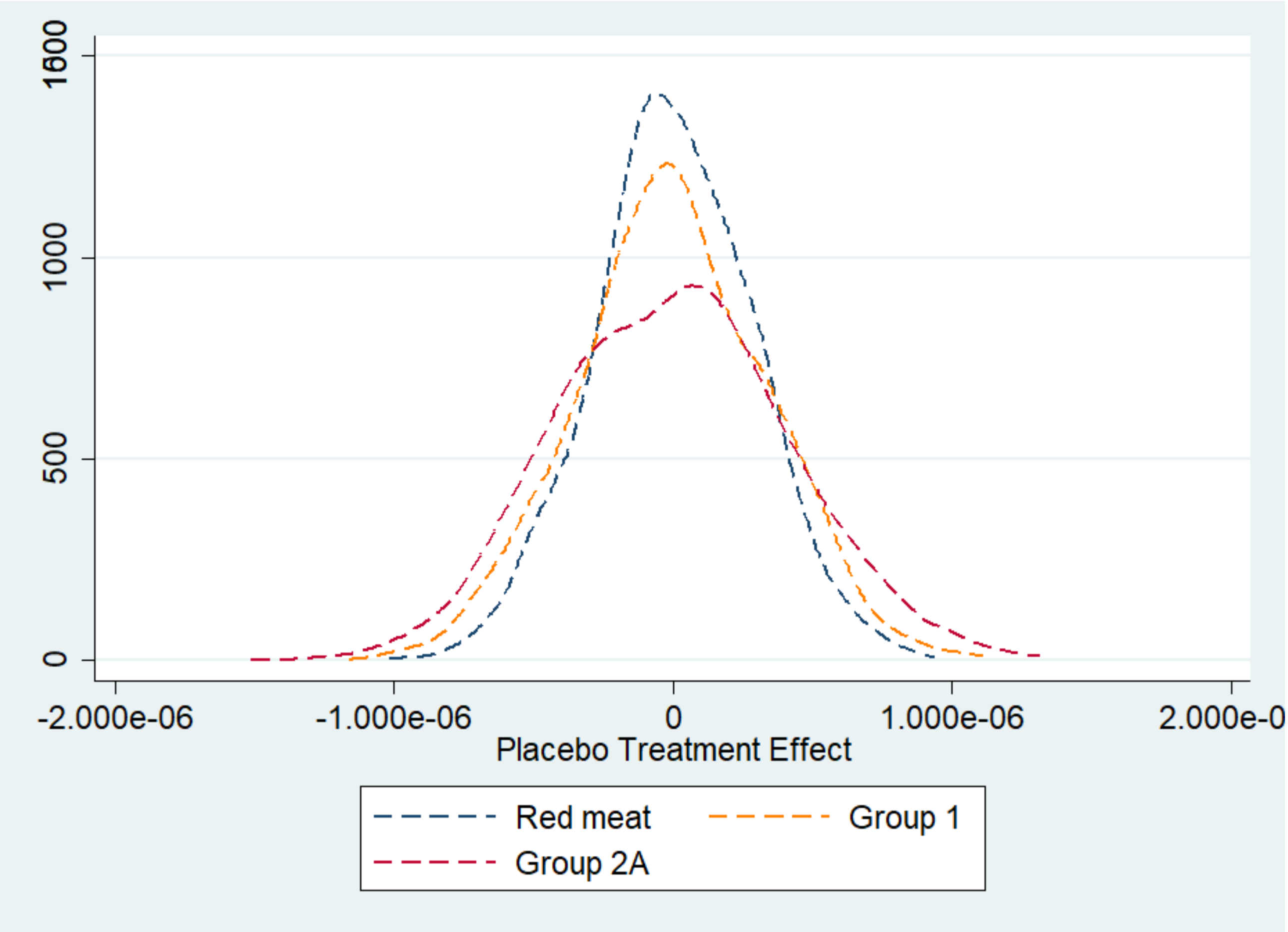


Figure 5. Kernel density estimates for placebo warnings



Appendix

Table A1. DiD estimates of the effect of the warning on meat expenditure: short-term effects (Bootstrapped standard errors)

	(1) Red	(2) Group 1	(3) Group 2A	(4) Red	(5) Group 1	(6) Group 2A
Treatment	-0.0491***	-0.0320***	-0.0742***	-0.0478***	-0.0313***	-0.0722***
	0.0053	*0.0081*	*0.0082*	*0.0050*	*0.0083*	*0.0089*
Food Exp.	0.0015***	0.0013***	0.0017***	0.0014***	0.0012***	0.0016***
	0.0000	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
HH size				0.0527***	0.0530***	0.0502***
				0.0017	*0.0028*	*0.0029*
High Educ.				-0.1049***	-0.0833***	-0.1388***
				0.0051	*0.0088*	*0.0076*
HH Age 35-65				0.0241***	-0.0025	0.0525***
				0.0074	*0.0100*	*0.0160*
HH Age>65				0.0446***	-0.0424***	0.1239***
				0.0110	*0.0135*	*0.0158*
Migrant				-0.1503***	-0.2204***	-0.1054***
				0.0118	*0.0169*	*0.0148*
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	30852	30852	30852	30852	30852	30852

OLS estimates coefficients of treatment effect of Equation (1). Full set of controls included. Bootstrapped standard errors clustered at month level based on 200 replications in *italics*. ***, **, * indicate significance at 1%, 5% and 10%, respectively.

CONCLUSIONS

This thesis investigated the two-way relationships between income and health and tried to disentangle the main mechanisms of transmissions of related inequalitiesIt analyses both distributional aspect of income - focusing on the top tail of income distribution and in particular on those individuals who get high earnings in the labour market (i.e. the so-called working super-rich)- and the socioeconomic gradient in health both in the top tails and in the rest of the distribution. The thesis consists of three chapters.

The first chapter focuses on the determinants that allow some individuals to receive extraordinary earnings in labour markets compared to their peers, i.e. CEO or the superstars of sport, music and cinema. Specifically, it empirically analyses the effects of performance, popularity and bargaining power on the earnings of the universe of football players of Italian Serie A. Disposing of original panel data allows to investigate the returns of performance, popularity and bargaining power while controlling for players and team unobserved heterogeneity. Moreover, I employ the Unconditional Quantile Regression approach developed by Firpo, Fortin and Lemieux (2009) to estimate the impact of a marginal change in the determinants of earnings on their entire distribution. This is relevant because players' earnings exhibit a large dispersion around the mean and investigating the role of the determinants is especially interesting at the top of the earnings distribution, where superstar effects should more clearly manifest. Main results show that all the aforementioned factors significantly affect the players' earnings. However, the analysis "beyond the mean" reveals that the role played by popularity increases at the top of earnings distribution being the main determinant of the "superstars". In particular, I find that popularity dominates all the other covariates at the top tail of the earnings distribution and reaches its peak at the 95th percentile of this distribution. An increase in a player's popularity is magnified by an earning premium of approximately 35% at its maximum point, according to my estimates. Conversely, the role of bargaining power reaches its peak around the 75th percentile, generating an earnings premium of approximately 14%, but it is not statistically associated with the earnings of players above this threshold. The effect of performances is, instead, more constant and becomes relatively less important after the 75th percentile. These results challenge the interpretations of extraordinary earnings based only on very talented workers who "win and take all".

As mentioned above, labour market seems a fertile ground for the escalation of contemporary society's extreme inequalities. Thus, understanding how the "working super-rich" respond to health shocks can be a key element of these dynamics. The second chapter of this thesis aims at providing evidence of the relationship between health shocks and labour market outcomes, focusing on those in the top tail of earnings distribution. Therefore, the dataset presented above has been enriched with data about the nature and the incidence of the injuries suffered by the universe of Serie A football players. In particular, in this chapter, I exploit traumatic injuries as exogenous variation in professional football players' health to provide estimates of the causal impact of a health shock on two main labour market outcomes: the annual net wages and the probability of changing the terms of the contract between the employer (the club) and the employee (the player).

The empirical approach employs panel fixed effects models combined with an IV strategy, which uses the average number of yellow cards received by the team as an instrument. Main results show that working super-rich are not immune to economics consequences of health shocks and can be summarised as follow. Firstly, health shocks have a strong negative effect on the wages of the professional football players of the Italian Serie A. In particular, having a 30-day injury reduces the wage of the following year by approximately 12%. Secondly, only a residual part of the negative effect could be explained through the reduced performance of the players after the injury. In fact, results suggest that the largest part of the coefficient is explained by a direct effect of the shock on the outcome. This result can be explained in a framework of human capital depreciation. The club has an incentive to offer a lower wage for precautionary reasons, supposing that players who experienced a severe injury might incur similar injuries in the future. Thus, the club insures itself against this risk by reducing the fixed share of the wage, which is independent by performances.

The third chapter analyses inequality and its multifaceted aspects by moving the focus from the top tail to the whole population. In fact, it investigates how people heterogeneously respond to public education policies delivering health information by using as case study the 2015 WHO warning about the carcinogenic effects of red meat consumption. Importantly, it exploits high frequency data, i.e. monthly data, about Italian households' expenditures to identify the effect of the warning in the long vs. short run, for which there is no previous evidence in the literature and to document the response of households differing due to educational levels and health awareness. In order to identify such effect, I employ a Difference-in-Difference model which exploits the strong seasonality patterns in meat consumption in Italy, mainly associated with culinary traditions in occasion of catholic holidays celebrations. Main findings can be summarised in three main points. First, I find that WHO's warning had a strongly significant impact on consumer's behaviour but only in the very short-term. In fact, I find that in the first month after the warning, consumers reduced their expenditure on red meat by around 5%, 3.2% and 7.4% of the average monthly expenditure in generic red meat, in carcinogenic meat and in probably carcinogenic meat, respectively. However, expenditures on red meat returned to pre-warning levels just two months after its release. Secondly, I find that only some subgroups, such as the more educated and health aware changed their eating behaviours in the long run, i.e. over a one year post-warning observational period. On the contrary, poor-educated and less health-aware households reduced the consumption in a less significant manner and only in the very short-term. Thirdly, I also find that these groups differ significantly with respect to the correct interpretation of the warning. More educated and health-aware households reduced especially the consumption of carcinogenic meat while their counterparts reduced mostly the consumption of relatively less dangerous meat.

These findings also have a number of implications for the successful design of health information policies. Firstly, the fact that the consumers - on average- only responded in the immediate short-term suggests that successful health information policies should be designed in a way that expose the individual to a constant flow of information. This is likely to be more beneficial compared to the "one-shot" warnings. Secondly, the misinterpretation of the warnings by some subgroups suggests the importance of re-designing the health warnings in a simpler and more accessible way. In this respect, a higher involvement of health care providers in dissemination of health warnings might be beneficial. Finally, the findings confirm the strategic role played by education for health. Other than to reduce the well-known health gap, my finding indirectly suggests also that education is able to increase the health returns on investments in health campaigns and health educational activities since the latter are misinterpreted by low educated individuals and produce only short-term effects among them. In a general equilibrium perspective, higher investments in education are then likely to bring both equity and efficiency gains to the health production process.