Workforce substitutability: The case of soccer players*

Santiago Velásquez †

Abstract

Quantitative analysis of labor production functions and input substitutability requires high-quality data on labor productivity that is sometimes difficult to observe. Professional sports matches present a highly suitable scenario for this type of analysis given the amount, frequency, and high-quality data usually collected. I study labor substitutability using the information on the top 5 most competitive men's soccer leagues in Europe. Focusing on injures that are not predictable, I use event study approaches to estimate the effect of the negative exogenous shock on the injured player's and his team's performance. The findings suggest that injuries do have long and medium-term effects on the injured player's performance. We also find that the extent to which a player's performance affects their teammates depends on the player and his team's types, with higher performers' being more consequential for low-performing teams.

Keywords: Event Study, Sports, Productivity

JEL Codes: D24, L83

*This paper was my undergrad thesis and I translated it to English for the purpose of the research assistant position application. The original version in Spanish can be accessed here.

†svelas45@eafit.edu.co

1 Introduction

In economic sciences, elements such as the supply and demand of goods and services are of frequent interest. In particular, regarding the firms, one is interested o how they produce and how the different productive factors interact in production. According to Rosen & Sanderson (2001), numerous essential elements of supply and demand can be seen in the sports activities labor markets, and many of these can and are investigated empirically.

Although it has been studied for decades, the literature related to the economics of sport (Rottenberg 1956, Neale 1964, Sloane 1971) is still in development due to the advancement of new statistical techniques, the increase in computing capacity and access to information, as well as the variety of applied theories and exciting hypotheses, which can be validated in sports contexts. In my case, leagues and soccer teams constitute a highly suitable scenario to test such hypotheses on classical economic theories such as the producer's theory given that, unlike other areas where production and the contribution of factors to it cannot always be measured directly, nor is it easily differentiable or quantifiable, in this context, it is possible to obtain performance measures that can represent both the production of the firm, measured as the results of the teams in their respective leagues and the contribution of the productive factors; specifically, the workforce, this being able to be quantified with performance variables of the players within each match. Moreover, the fact that such production by the firm and the workforce's productivity can be measured constantly over relatively short intervals of time allows the creation of panel-type databases, which represents a substantial advantage for the use of econometric techniques.

Many previous studies exploiting sports data have focused on labor demand, factor substitution, wages (Caruso et al. 2017), etc., as documented by Rosen & Sanderson (2001) in their study. While some of the most recent works have focused their attention on testing hypotheses related to other areas of economics, such as behavioral economics Gauriot & Page (2019), Bryson et al. (2020), conversely, my work focuses on identifying the contributions of productive factors to production using exogenous events that allow the effect of being causally quantified.

In the particular case of this research, the approach is to delve into the measurement of the contribution of productive factors, specifically workforce, to the production function of firms, using data from soccer games for the five main European Leagues: Bundesliga (Germany), La Liga (Spain), Ligue One (France), Premier League (England) and Serie A (Italy), with game-by-game information for each player, from the 2014-2015 season to 2018–2019.

Taking advantage of the set of unexpected events, defined by the players' injuries during competition, it is possible to identify how these injuries affect the contribution of a player in the upcoming matches and how the absence of this production factor later affects the performance and production of the firm. Allowing me to answer questions such as: what is the first-order effect of these injuries on the players' participation in the game? And is it possible for the firm to completely compensate/supply the impact generated by the absence of a player?

The estimations of this study allow me to identify parameters of the production function that are difficult to capture in other contexts. Understand the sensitivity of a firm's production system to changes in production factors and contemplate possible changes in the production function according to the productivity level of the affected individual or his team, being understood as the cross-elasticities between these factors.

This study constitutes a contribution of general interest, particularly for microeconomics and organizational theory, being also focused on people from various areas of knowledge who are interested in knowing a little more about findings related to sports.

The rest of this work divides as follows: Section 2 includes a review of the theoretical framework of the economics of soccer and the contribution of labor to production in team contexts, especially in sports. Section 3 presents descriptive statistics of the data used in this study, while section 4 discusses the methodology used. Section 5 presents the results and their heterogeneous effects, and then I conclude and describe the next steps finally.

2 Theoretical Framework

Soccer is the most played sport worldwide (Dvorak et al. 2004). According to Efe (2020) in 2018, the five main European leagues had revenue between 1,700 and 5,400 million euros, thus representing around 75% of the total income of all the leagues of the 55 countries belonging to the Union of European Football Associations (UEFA). In turn, the development of this activity has impacts on the economy of nations; for example, according to Ernst&Young-LLP (2019) during 2018, the Premier League generated 3.3 billion in taxes, contributing 7.6 billion to the GDP of the United Kingdom and maintained around 12,000 direct jobs and 100,000 indirect jobs. The preceding is an example of this sport's social and economic importance worldwide, which is an additional motivation to focus on it for academic studies.

As in most highly competitive sports, physical problems and injuries play a crucial role in the performance of athletes. The effect of injuries on the performance of soccer players has been previously evaluated by several academic studies focused mainly on the area of medicine (Stubbe et al. 2015, Zuke et al. 2018)); these studies have found significant effects on the number of minutes played and different variables of performance and return to competition, even though the latter has an extensive definition Zuke et al. (2018). Despite these approaches from sports medicine, there is still a lack of empirical evidence that studies this type of event and its implications from an economic point of view.

Given the nature of the study, it is essential to divide the possible approaches to the subject into two large areas of research and their intersection. In the first place, the study by Rosen & Sanderson (2001) provides state of the art at the time of its publication of previous works around sports economics. It analyzes in detail the economic problems that can appear in this context. Considering that "sports is an exceptional forum for empirical economic analysis" (Rosen & Sanderson 2001), one can review studies on labor demand, factor substitution, wages, the marginal product of labor, wage discrimination, labor supply, as well as various topics of industrial organization, negotiation and competition. Likewise, Scully (1995) points out that the results at the team level can be associated with measures of their performance, and in turn, the contributions to these results are part of the production function and can be estimated as the marginal product of said function. On the other hand, some other studies adopt an empirical context more from the point of view of sports associations (leagues) and entertainment business than that of teams as profit-maximizing firms, which represents a significant difference from my work.

More related to my research, the study by Bonhomme (2020) proposes a methodology to estimate individual contribution in team contexts where only the aggregate result is observed, beyond the fact that in my case, I can also measure individual performance, this study presents a detailed conceptual framework of the theoretical problem with which I am dealing.

According to Bonhomme (2020), one is interested in certain primary relationships between workers and teams. First, like Arcidiacono et al. (2017)), he considers the fact that workers are heterogeneous in productivity, which later affects the team's total performance. Therefore, it is of the utmost importance to have data with a network-type structure or employee-employer with mobility. In this way, it is possible to show the workers' performance in different teams, measuring which contributes the most. Second, he considers the relations of the organization of workers between teams and the complementarity of the workforce in this conceptual framework. Finally, he considers other factors that fall into the model error as noise in the estimation.

Broadly speaking, Bonhomme (2020) uses two types of production functions (additive and non-linear production). Using two key transversal assumptions, such as the network exogeneity assumption, which establishes that the team formation process is independent to team shocks conditional on worker types, and a serial independence assumption that assumes independence on shocks of teams repeatedly collaborating over time. An advantage of my identification strategy is that I do not require applying these assumptions, which can be quite restrictive in soccer.

Considering the above, it is possible to decompose the variation of the output and assign which part corresponds to each of the relationships mentioned previously.

Finally, it is worth mentioning studies such as those by Arcidiacono et al. (2017) and Devereux (2018), who also try to measure individual contributions to team results but without considering some innovations proposed in my study, such as the use of unexpected exogenous shocks to quantify the effects causally and the fact that the assumptions of serial independence and exogeneity of the network are not required. Both studies are framed in the sports context; the first uses data on basketball and the second on tennis. Thus, the study by Arcidiacono et al. (2017)) focuses on estimating the importance of players in the performance of their teammates, developing a model that allows the quantification of this effect through an algorithm to compute the parameters proposed in the study. On the other hand, Devereux (2018) finds that around 50% of the variation in results at the team level is explained by team skills rather than individual skills, exploiting the fact that players compete both individually and in pairs in tennis.

It is essential to highlight that the studies mentioned in this last section have been developed recently and, in the same way, make use of new methodologies and sophisticated mathematical models that are at the frontier of scientific development. Likewise, they have more recent data, which is evidence of the results' current legitimacy. Therefore, these mentioned reasons give strength to the empirical validity of my exercise.

3 Data

The data available contains match-by-match information for each player of all the teams of the five main soccer leagues in Europe – Premier League (England), La Liga (Spain), Serie A (Italy), Bundesliga (Germany), and Ligue One (France) – with an analysis period of 5 seasons, from the 2014-2015 season to the 2018-2019 season, these having been extracted automatically using web scraping methods.

Overall, there is information for 136 teams, 6,145 players, and 9,130 games, as well as multiple productivity variables at different aggregation levels, such as performance throughout the season or during a specific game: of both the team and each of the players. To address my empirical question, it is crucial to have information for all the players regardless of whether they had been playing in the matches or if, on the contrary, they were substitutes or not called up, since the latter constitute the team's options to replace the usual players if they are affected by injury. Table 1 contains descriptive statistics for the main variables of interest; panel A contains information for the entire sample, while panel B only restricts the matches in which the players had action.

Table 1: Variable descriptive statistics of performance. Own elaboration.

	Panel A. Full sample				Panel B. Participating players			
	Individual Variables							
	Mean	Standard dev.	P1	P99	Mean	Standard dev.	P1	P99
Assistances	0.0402	0.2126	0	1.000	0.0825	0.2986	0	1.000
Expected Assistances	0.0363	0.1244	0	0.646	0.0743	0.1700	0	0.808
Shots	0.4629	1.0444	0	5.000	0.9487	1.3321	0	6.000
Goals	0.0506	0.2493	0	1.000	0.1037	0.3492	0	2.000
Expected Goals	0.0505	0.1725	0	0.883	0.1035	0.2356	0	1.137
Not Penalty Goals	0.0461	0.2352	0	1.000	0.0944	0.3299	0	1.000
Minutes played	35.283	41.165	0	90.000	72.319	28.194	1	90.000
NPGX	0.0460	0.1547	0	0.789	0.0943	0.2110	0	1.012
Key Passes	0.3501	0.8424	0	4.000	0.7175	1.0913	0	5.000
Yellow Cards	0.0733	0.2607	0	1.000	0.1503	0.3574	0	1.000
Red Cards	0.0018	0.0420	0	0.000	0.0036	0.0601	0	0.000
XGBuildUp	0.0794	0.2015	0	0.973	0.1627	0.2640	0	1.199
XGChain	0.1344	0.2953	0	1.386	0.2756	0.3739	0	1.685
	Team Variables							
Expected Assistances	0.9226	0.6735	0.046	3.118	0.9479	0.6869	0.050	3.165
Goal Difference	0.0586	1.8770	-5.000	5.000	0.1147	1.8897	-4.000	5.000
Shots	11.980	5.1836	3.000	27.000	12.1674	5.2170	3.000	27.000
Goals Scored	1.3986	1.2759	0.000	5.000	1.4343	1.2971	0.000	5.000
Goals Against	1.3400	1.2466	0.000	5.000	1.3196	1.2388	0.000	5.000
Expected Goals	1.2986	0.8804	0.093	4.032	1.3319	0.8958	0.098	4.123
NPGX	1.1895	0.8133	0.090	3.793	1.2211	0.8283	0.095	3.857
Key passes	8.9640	4.2630	1.000	21.000	9.1130	4.3009	2.000	21.000
Probability of Defeat	0.3636	0.4810	0.000	1.000	0.3530	0.4779	0.000	1.000
Probability of Tie	0.2487	0.4323	0.000	1.000	0.2463	0.4308	0.000	1.000
Probability of Victory	0.3877	0.4872	0.000	1.000	0.4008	0.4901	0.000	1.000
XGBuildUp	2.0397	2.1487	0.000	9.993	2.1144	2.2142	0.000	10.381
XGChain	3.4531	3.0573	0.116	14.229	3.5699	3.1445	0.119	14.723

Nota:Panel A, N=688.412 Observations. Panel B, N=335.864 Observations.

As can be seen, there are traditional performance metrics such as goals, assists, shots, or key passes. One of the potential restrictions associated with soccer is the fact that in this sport, the number of goals per match is generally low compared to others such as basketball (Gaviria (2000), Matano et al. (2018)). To overcome this problem, in my study, I also have new indicators that allow me to measure with greater precision and variability the performance of a player on the field, such as the probability of scoring or making important plays within a game. The "Expected Goals" and "Expected Assists" are within these new metrics, which are obtained through predictive models. These indicate the probability of success of the play from variables such as distance of the shot, angle, type of play (Free kick, individual play, corner kick) part of the body used for the shot (leg, head, among others) without these measures being conditional on the final result of the event as described in Bundesliga (2018).

On the other hand, there is information about players' status in each match, including the position on the field and any different possible situations. Including whether the player was injured and, in turn, the type of injury for which he was absent. Given the heterogeneity in the players and the long period that I am analyzing, it is possible to find various types of injuries. So for this analysis, all injuries that had occurred less than 80 times are grouped into similar categories. Table 2 gives an idea of the types of injuries found in the data and their frequency. To avoid double-counting data due to correlated or simultaneous injuries, my main estimates include only injuries that another injury had not preceded in the last eight months, which is why table 2 presents the frequencies for both total injuries and the number of events once the 8-month restriction is imposed. [1].

To guarantee both the homogeneity of the data and the matches used, only the games corresponding to local leagues will be used since these have a similar system of a table of positions formed with the result of the matches all against all. And a series of incentives based on the team's standing by the end of the tournament, with prizes, qualifications for international tournaments, and relegation. The preceding generates that some players are

[1] The main results are robust to the inclusion of all injuries

Table 2: Number of injuries suffered by type. Own elaboration.

Injury	Without restriction	8-month restriction
Abdominal/Intestinal	127	34
Crash	225	79
Surgery	120	48
Cruciate ligament	141	58
Muscle tear	235	59
Fatigue	105	27
Fracture	295	103
Hits	198	55
Flu/Allergies	330	104
Wounds	87	36
Influenza	147	28
Hamstring injury	875	263
Unknown injury	448	165
Shoulder injury	132	143
Muscle injury	607	45
Calf injury	214	192
Foot injury	118	43
Knee injury	467	47
Ankle injury	380	131
Injury/Problems	860	262
Ligaments	474	155
Muscular	400	124
Bone (No fracture)	320	110
Other	258	73
Adductor problem	362	103
Back problem	138	37
Muscle problem	544	135
Thigh problem	435	109
Calf problem	130	23
Knee problems	256	66
Achilles tendon	116	34
Tendons	170	50
Total	9714	2941

censored when they stop playing for such leagues; however, the results are robust to the exclusion of these players.

4 Methodology

Given the availability of data and the different performance variables mentioned above, it is possible to use the measurements of each match for the players to correctly contrast the type of theories related to performance in team contexts. And the contribution of labor to the total output of such teams or economic units. By making use of exogenous shocks that happen to individuals in different periods, it is possible to adjust the available data to perform an event study type estimation, which allows computing the effect of said shocks on the variables of interest and has been extensively studied previously (Freyaldenhoven et al. 2019, Roth 2018).

In our case, exogenous shocks are defined by the set of injuries that occur to players during the season, and I am willing to identify their effect on performance variables both at the individual level and the aggregate results of the team to which the player affected by the injury belongs.

According to Abraham & Sun (2018), due to the existence of panel-type databases, it is common in research to estimate treatment effects using temporal and individual ("two-way") fixed effects. Similarly, one is interested in assessing the dynamics underlying these treatment effects, which is often done by estimating an equation such as the following:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{\tau} \mu_i 1[t - E_i = \tau] + v_{i,t} \tag{1}$$

"Where $Y_{i,t}$ is a variable of interest for a unit i in period t, E_T is the period where unit i was initially treated and α_i and λ_t are the fixed effects of individual and time... The relative time $\tau = t - E_i$ included in (1) covers most of the relative periods but may exclude others." (Abraham & Sun 2018). In turn, this methodology has three key identification assumptions that must be understood to make a correct estimate:

- Assumption 1: Parallel trends in the variables for the periods prior to the shock.
- Assumption 2: There is no anticipatory behavior prior to treatment.
- Assumption 3: Homogeneity in the treatment effect. [2]

The first two assumptions are similar to the traditional assumption of parallel trends of the difference-in-differences model, in which it is assumed that in the absence of the event, the differences between the trends of the treated group and the control group would have remained constant. In contrast, assumption three refers to each treated cohort experiencing the effects of the treatment in the same way.

[2]Recent studies such as Abraham & Sun (2018), Roth (2018), Freyaldenhoven et al. (2019) have developed new methodologies to deal with violations of these assumptions and achieve better identification. For the purposes of this study, the discussion on the validity of this assumption will be addressed later in the text.

Considering the above, this methodology presents appropriate characteristics for its use in empirical studies [3]; therefore, in my study, an equation of the following form was estimated:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{k=-6}^{6} \beta_k D_{i,t}^k + \phi_{t<-6} + \phi_{t>6} + v_{i,t} \tag{2}$$

Where α_i y λ_t are the fixed effects of individual and time and the coefficients of ϕ are associated with variables that indicate the periods before -6 and after 6 concerning the period where the event occurs, defined as the match where the injury occurs or the team's next game if the injury occurs outside of a game. To comply with the assumptions above, injuries that were not preceded by another injury in the previous eight months were considered. Also, given the possibility that a player could experience multiple injuries that meet the above restriction, estimates were made using player-injury level data and including player fixed effects, as mentioned in the equation, and tournament week and season fixed effects identify significant temporal variation in this context.

Additionally, given the heterogeneity of injuries that individuals may face in the sporting context, it is challenging to empirically argue that all of these comply with the assumptions above (mainly assumption 2). Since some of these are dynamically affected by realizations of the previous periods, i.e., an increase in the number of minutes played or a muscular overload can affect the players' performance long before the period where treatment is evidenced in the data. To deal with this problem in the cleanest way possible, a data-based algorithm was designed that identifies, by extracting a random sample, the injuries that are most likely to be predicted in the data thus excluding them from the analysis.

In other words, for each type of existing injury, an event study is estimated in a random sample of 20 percent of individuals with this injury (training sample). A hypothesis test is performed on the probability that the associated coefficients to the periods before treatment are jointly different from zero, and the p-value associated with the test is saved. Finally, for

[3] Vease Jacobson et al. (1993)

the remaining 80 percent, the equation is estimated again without considering the types of injuries with a p-value less than 0.2 in the exercises of the training sample.

This way, I ensure that the estimates of my model are not affected since I am left only with the types of injuries that are not very predictable according to the data. In addition, due to the use of random sampling and the design of the algorithm wholly data-driven, I make sure that I am deliberately selecting the sample to a minimum, thus avoiding bias, given the orthogonality of the randomly chosen data.

Finally, as will be seen in the next section, it is possible to estimate heterogeneous effects by making use of the possibility of creating, through the data, categories of higher/lower productivity both at the team level and at the individual level, seeing wherein the distribution were located just before the period of injury and partitioning by the 75th percentile. With this, I can empirically evidence in which class of players, teams, or player-team combination the effects are mainly concentrated by estimating an equation like the following:

$$Y_{i,t} = \alpha_i + \lambda_t + X'_{i,t}\phi + \sum_{\pi}[\beta_{pre}D^{pre}_{i,t}I_\pi + \beta_{post}D^{post}_{i,t}I_\pi] + \phi_{t<-6} + \phi_{t>6} + v_{i,t} \tag{3}$$

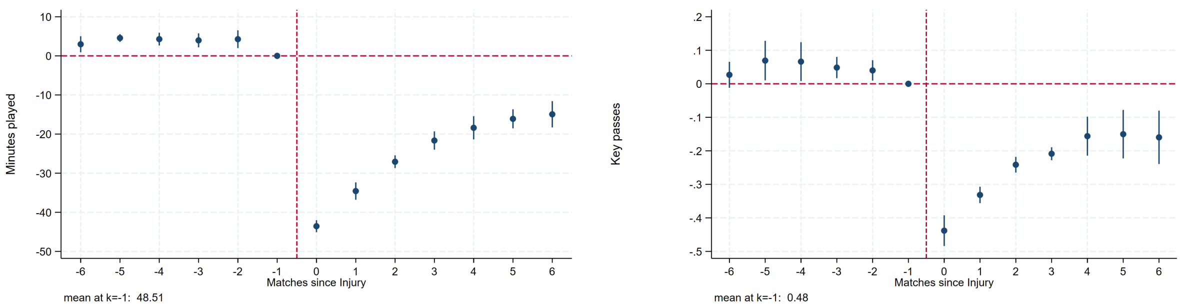
Where all the variables are defined in the same way as in 2 and π represents the number of groups among which the effect is to be seen, and I_π is an indicator variable that takes the value of one of the individuals i belongs to group π or zero in another case. Unlike the event study regression where the dynamics of the coefficient is studied, on this occasion, I group the coefficients of the periods before and after the shock to ensure that I have enough power in the data to estimate the effect of each group by not having to assess for many parameters.

5 Results

Figure 1a 1a shows the coefficients associated with the exercise before applying the data-based algorithm to select the less predictable types of injuries, using the variables minutes played and key passes as measures of player participation in the match as outcomes. As could be expected, not all injuries are unpredictable; on the contrary, when considering all types of injuries in the database, I see significant coefficients associated with the parameters of the periods before the event. These "pre-trends" indicate that, compared to the period immediately before the shock, in the periods before this, the players had been having more minutes per game, as evidenced in graph 1a, which may be evidence of anticipatory behavior.

In turn, graph 1b shows me how, in addition to the existence of significant coefficients in the periods before the event, the key passes variable also indicates a trend, that is, in the periods before the reference period, the number of key passes made by a player in each game was significantly higher. However, they had been decreasing in quantity, which provides another type of evidence that some of the types of injuries in the database are not entirely unpredictable, but instead affect the player partially in previous periods until finally, he must be absent due to the injury.

Figure 1: Estimation without applying the algorithm



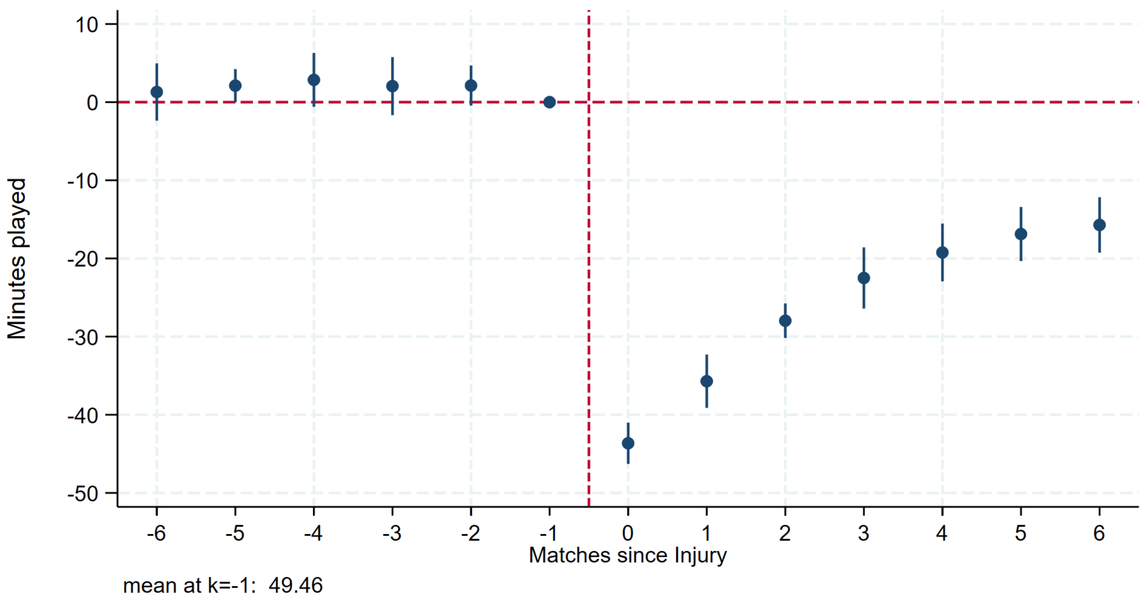
(a) Minutes played (b) Key passes

Note: This Figure shows the event study coefficients associated with the estimation of equation (2) using all injuries. The results show the potential presence of anticipatory behaviors. Estimates include individual, match-week-season fixed effects. Cluster standard errors at the individual, game week-season level.

Following the above, it is vital to ensure that the developed algorithm manages to classify the types of injuries in such a way that the less predictable injuries are separated from those that influence the performance of the players in previous periods to comply with the assumptions of the method and make valid causal estimates. Figure 2 shows the effect of injuries on the number of minutes played when applying the data-driven classification algorithm. In this, I can see how, in comparison with figure 1a, the coefficients associated with the periods before the shock are not statistically significant at 5%, nor is there an evident trend in the punctual estimators of such parameters.

Nonetheless, the periods after the event do turn out to be highly significant. For example, in period 0, the players are present on average 42 minutes less, which, compared to the average of the base period, is a relative reduction of 85% minutes played, being this effect persistent during the next six games, although the punctual estimator decreases as the period of the shock recedes. In other words, the players continue to play fewer minutes, although this decrease is slighter and less because as the games go by, it is likely that the players will recover from the injury.

Figure 2: Effect of injury on the number of minutes played



Note: This Figure shows the event study coefficients associated with the estimation of equation (2) using the lesions selected with the classification algorithm. Estimates include individual, match-week-season fixed effects. Cluster standard errors at the individual, game week-season level.

Just as the number of minutes played is affected by injuries, the performances and participation of these players in the matches also suffer an effect. Figure 3contains the results of the impact of injuries on different types of individual performance variables. I can see how, regardless of the measure chosen, there is evidence that injury generates a negative and significant impact on player productivity, measured in this case as performance within matches. That is, the productive factor of labor effectively suffers a negative shock that is evident in its productivity.

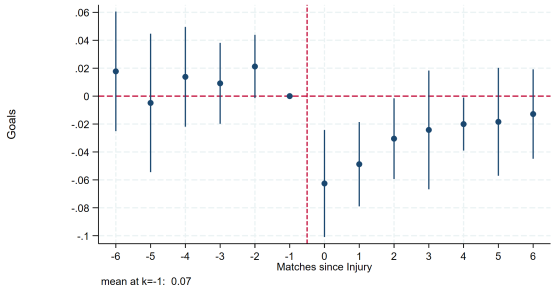
For instance, I see how the number of real goals (Figure 3a), and expected goals (Figure 3b), has an immediate relative decrease of approximately 90%. In turn, I see how the estimate for expected goals is more precise than its counterpart in terms of standard error. This is because, as mentioned above, variables such as goals have little variability in low-scoring sports such as soccer; thus, the effect in expected goals is persistent for more periods, although in the same way, after approximately six games, the impact stops being statistically different from zero.

Also, the effect on other variables of individual performance presents a similar dynamic, in which the period of the shock presents the impact of greater magnitude, and this decreases as the matches progress. It should be noted that, unlike figure 1b, in this case, none of the variables shows a trend in the periods before the shock.

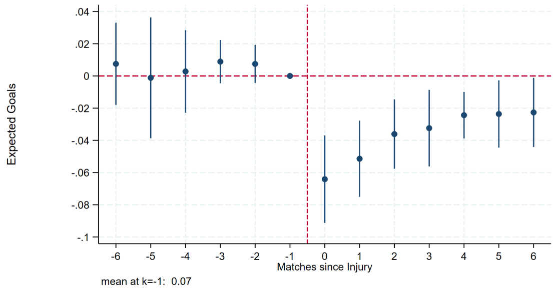
Another essential point to mention is the effect on the XGBuildUp variable. This is an innovative variable in football metrics as it quantifies the total expected goals of all the possessions in which the player was involved in the game, discounting key passes and shots during the 90 minutes, i.e., this variable gives me an aggregate measure of the player's performance, with the *compliers* being the offensive players since most of these variables are measures of offensive performance.

Considering the dynamics presented by the previous results, being free of "pre-trends," I will continue analyzing results of estimates like the difference-in-differences model, where the coefficient presented is interpreted as the average of the effect in the six periods follow-

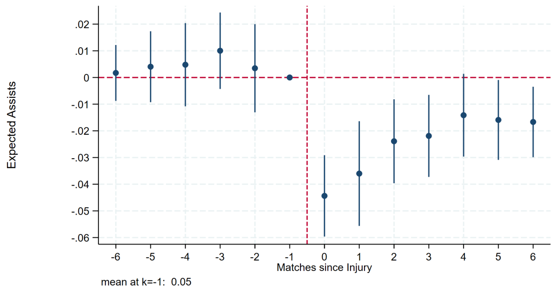
Figure 3: Effect of injury on individual variables



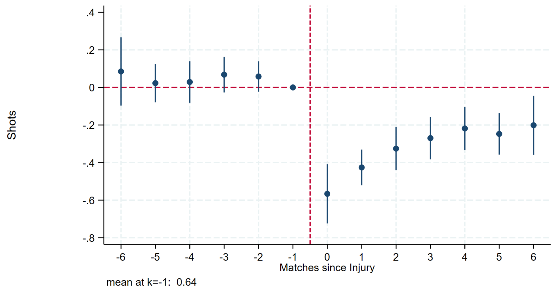
(a) Goals



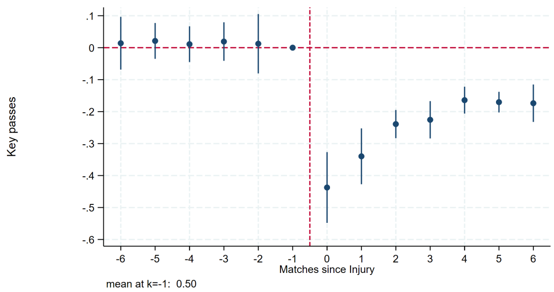
(b) Expected goals



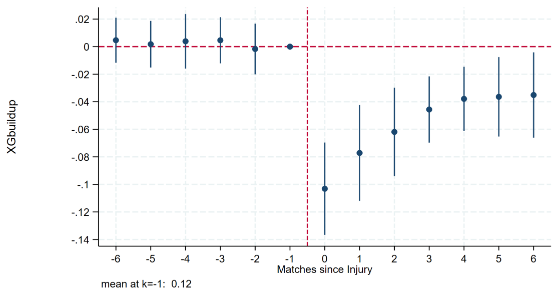
(c) Expected Assistances



(d) Shots



(e) Key Passes



(f) XGBuildUp

Note: This Figure shows the event study coefficients associated with the estimation of equation (2) using only the least predictable injuries according to the data-driven algorithm. There is evidence of a decrease in player productivity measured using different performance variables. Estimates include individual, match-week-season fixed effects. Cluster standard errors at the individual, game week-season level.

ing the event [4].

Table 3 summarizes the evidence of the effect of the shock on the individual results, similar to the graphs. I see a negative and significant impact on all the variables. For exam-

[4]The results for the rest of the variables of interest can be found in Appendix A.

Table 3: Effect of injury on individual variables

	(1) Minutes Played	(2) Goals	(3) Expected Goals	(4) Expected Assistances	(5) Shots	(6) Key Passes	(7) XGbuildup
Pre	2.087*	0.0114	0.00509	0.00480	0.0526	0.0155	0.00264
	(0.971)	(0.00829)	(0.00579)	(0.00463)	(0.0317)	(0.0157)	(0.00463)
Post	-25.95***	-0.0310**	-0.0364***	-0.0247***	-0.322***	-0.250***	-0.0568***
	(0.816)	(0.00997)	(0.00618)	(0.00438)	(0.0346)	(0.0120)	(0.00973)
Constant	43.29***	0.0548***	0.0600***	0.0421***	0.549***	0.424***	0.0959***
	(1.019)	(0.00820)	(0.00469)	(0.00417)	(0.0293)	(0.0183)	(0.00766)
Observations	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.339	0.124	0.225	0.142	0.297	0.235	0.159
Mean at t-1	49.46	0.0700	0.0700	0.0500	0.640	0.500	0.120

Note: Estimates include individual, match-week-season fixed effects. Cluster standard errors at the individual, game week-season level. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

ple, there is a 52.4% relative reduction in the average number of minutes played in the next six games, a 44% reduction in the probability of scoring a goal, and a 52% reduction in the expected goals. Similarly, the rest of the variables show decreases of around 50%. These results show how the set of selected injuries presents behavior that allows me to identify the effect of these events on individual performance causally.

Now, a relevant question in the context of a production system is how these drops in the productivity of the players can affect the firm's total productivity. I ran regressions where the dependent variable is team performance to answer this question. Table 4 shows the effect of the shock on the variables at the team level. Contrary to the above, I see how the coefficients associated with the group performance variables, despite also being negative, are only marginally or not significant. That is, on average, individual injuries do not seem to have a statistically significant or large effect on the aggregate playmaking variables measured in number of key passes, *NPXG*, number of shots, goals, and expected assists; or more complex measures like *XGBuildUp* or *XGChain*.

Given the evidence of the figures and tables presented so far I cannot reject the hypothesis that, even though an exogenous shock generated by the absence of workforce affects the productivity of the productive factor, the team can respond optimally to this event to compensate in some way for such a decrease and that the aggregate output is not affected.

Table 4: Effect of injury on team variables

	(1) Key passes	(2) XGChain	(3) NPXG	(4) Shots	(5) Expected Assistances	(6) Expected Goals	(7) XGBuildUp
Pre	-0.172	-0.0417	-0.00596	-0.243	-0.00832	0.00160	-0.0339
	(0.117)	(0.0756)	(0.00719)	(0.152)	(0.00859)	(0.00916)	(0.0666)
Post	-0.223	-0.103	-0.0308*	-0.331*	-0.0182	-0.0207	-0.0772
	(0.115)	(0.0571)	(0.0115)	(0.147)	(0.0129)	(0.0153)	(0.0500)
Constant	9.052***	3.472***	1.192***	12.13***	0.920***	1.293***	2.060***
	(0.0854)	(0.0773)	(0.0170)	(0.115)	(0.0125)	(0.0190)	(0.0604)
Observations	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.175	0.218	0.183	0.171	0.178	0.177	0.209
Mean at t-1	9.270	3.690	1.240	12.41	0.960	1.350	2.200

Note: Estimates include individual, matchweek-season fixed effects. Cluster standard errors at team, game week-season level. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

The teams manage to replace the contribution of the injured players so that the team performance, on average, is not altered. However, my lack of power to reject this hypothesis may be due to the possibility of mixing heterogeneous effects that, when studied individually, could shed more light on the impact of these exogenous shocks on individual productivity, a heterogeneity that I intend to exploit in the next section.

5.1 Heterogeneous Effects

This section includes the results of heterogeneous effects using the double interaction between teams with high or low productivity and players with high or low productivity. To carry out such a classification, the distribution of the variable *XGChain* cumulative until the date of the injury was considered individually and as a team. Those above the 75th percentile were classified as high productivity.[5] The importance of this variable, which adds the expected goals of all the plays in which the player was involved, is that it has preferable characteristics in my context, such as the fact that it has more significant variability for players in all positions on the field, which allows better identification of the level of productivity.

Regarding the individual productivity results, Table 5 shows how the effect of the injury

[5]Given the absence of clearly defensive court indicators, most of the variables analyzed in the study mainly consider the offensive aspect of the players' performance.

Table 5: Heterogeneous effects of injury on individual variables

	(1) Minutes Played	(2) Goals	(3) Expected Goals	(4) Expected Assistances	(5) Shots	(6) Key Passes	(7) XGbuildup
(a) Low P. - Low T.	-26.49***	-0.0167*	-0.0243***	-0.0187**	-0.241***	-0.192***	-0.0436***
	(1.391)	(0.00782)	(0.00430)	(0.00497)	(0.0311)	(0.0168)	(0.00697)
(b) Low P. - High T.	-19.13***	-0.0200	-0.0225*	-0.0109	-0.180*	-0.123***	-0.0439***
	(1.207)	(0.0107)	(0.00935)	(0.00514)	(0.0696)	(0.0256)	(0.00864)
(c) High P. - Low T.	-28.27***	-0.0487**	-0.0535***	-0.0288***	-0.450***	-0.323***	-0.0611***
	(1.135)	(0.0129)	(0.00726)	(0.00447)	(0.0394)	(0.0226)	(0.0108)
(d) High P. - High T.	-26.12***	-0.0448**	-0.0481***	-0.0392***	-0.414***	-0.348***	-0.0842***
	(1.037)	(0.0141)	(0.00938)	(0.00513)	(0.0376)	(0.0272)	(0.0129)
Constant	42.89***	0.0536***	0.0590***	0.0413***	0.538***	0.416***	0.0944***
	(1.066)	(0.00852)	(0.00490)	(0.00407)	(0.0312)	(0.0183)	(0.00776)
Observations	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.340	0.124	0.225	0.142	0.298	0.236	0.159
Mean at t-1	49.46	0.0700	0.0700	0.0500	0.640	0.500	0.120
Mean (a)	44.39	0.0200	0.0400	0.0200	0.370	0.270	0.0700
Mean (b)	33.81	0.0500	0.0400	0.0200	0.270	0.140	0.0700
Mean (c)	56.14	0.0700	0.0900	0.0700	0.840	0.730	0.130
Mean (d)	60.36	0.160	0.130	0.100	1.130	0.860	0.220

Note: Estimates include individual, game week-season fixed effects. Cluster standard errors at the individual, game week-season level. (a) low productivity player - low productivity team. (b) low productivity player - high productivity team. (c) High productivity player - low productivity team. (d) high productivity player - high productivity team. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

is maintained in most cases on the four analysis groups, showing relatively similar magnitudes of the coefficients associated with the impact. However, the effect on goals and shots ceases to be significant for the group comprising low productivity players in high productivity teams. And that the effect on expected assists and key passes, even though they continue to be significant, is only marginally so, probably due to a lower precision in the estimation.

Table 6, on the other hand, presents the results at the team level, where I can observe the effects on the aggregate performance variables. The importance of this table lies in the fact that empirically, it is seen that some teams are affected by the absence of a particular type of player. That could mean that maybe the results of the estimates of aggregate variables, previously presented, are statistically insignificant mainly due to the non-identification of a relevant group or, what is similar, the inclusion of particular noise in the estimates that do not allow to correctly validate the hypothesis of the effect of an individual shock on the aggregate performance. In turn, a player of a certain skill level can have a different impact depending on the type of players around him, i.e., the quality of the team to which he be-

Table 6: Heterogeneous effects of injury on team-level variables

	(1) Key passes	(2) XGChain	(3) NPXG	(4) Shots	(5) Expected Assistances	(6) Expected Goals	(7) XGBuildUp
(a) Low P. - Low T.	-0.113	-0.0483	-0.0160	-0.206	-0.00453	-0.000947	-0.0420
	(0.120)	(0.0607)	(0.0154)	(0.127)	(0.0121)	(0.0186)	(0.0554)
(b) Low P. - High T.	-0.150	-0.166	-0.0265	-0.232	-0.0170	-0.0136	-0.130
	(0.211)	(0.102)	(0.0397)	(0.310)	(0.0349)	(0.0410)	(0.0630)
(c) High P. - Low T.	-0.614***	-0.270**	-0.0747**	-0.754***	-0.0571**	-0.0756**	-0.179**
	(0.112)	(0.0779)	(0.0172)	(0.112)	(0.0162)	(0.0222)	(0.0626)
(d) High P. - High T.	-0.0382	0.0114	-0.0125	-0.152	-0.00159	-0.00126	-0.00173
	(0.210)	(0.0905)	(0.0306)	(0.286)	(0.0264)	(0.0272)	(0.0730)
Constant	9.036***	3.470***	1.191***	12.12***	0.919***	1.291***	2.060***
	(0.0788)	(0.0755)	(0.0162)	(0.113)	(0.0120)	(0.0182)	(0.0592)
Observations	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.175	0.218	0.183	0.172	0.178	0.177	0.209
Mean at t-1	9.270	3.690	1.240	12.41	0.960	1.350	2.200
Mean (a)	8.190	2.840	1.050	11.21	0.790	1.130	1.600
Mean (b)	10.15	4.280	1.360	13.43	1.070	1.510	2.650
Mean (c)	8.410	3.120	1.100	11.27	0.860	1.170	1.820
Mean (d)	11.79	5.600	1.690	15.39	1.330	1.850	3.520

Note: Estimates include individual, match week-season fixed effects. Cluster standard errors at team level, game week-season level. (a) low productivity player - low productivity team. (b) low productivity player - high productivity team. (c) High productivity player - low productivity team. (d) high productivity player - high productivity team. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

longs; therefore, it is essential to control at the same time the two variations margins as it happens in table 6

The results of these estimations indicate that, for the group of high productivity players in low productivity teams, the variables of key passes, *XGChain*, expected non-penalty goals (*NPXG*), shots, expected assists, expected goals, and *XGBuildUp* become significant, showing a relative decrease of between 6% and 8%. This is the only group of the four that shows substantial effects at 5%.

In addition, this result turns out to be highly intuitive since it is not surprising that the teams with the lowest productivity are the ones that depend on the most on their players, even more so if it is a player who at the time of his injury had a high level of productivity, which in turn is in line with the popular beliefs of soccer fans [6]

[6]The effects on the probability of defeat, draw or victory were also estimated where, perhaps due to the low variability of these measures, I did not find significant effects. (See Appendix A).

Therefore, this differentiation of the teams and players suggests that there are specific characteristics under which it is not always possible to supply a shock fully and optimally in the productivity of the factors of production. Moreover, these shocks decrease the team's ability to team play. Finally, even though the evidence on the probability of defeat, draw, or victory is not statistically significant, I see that the signs associated with these coefficients are intuitive for the group in question.

Table 7 presents an estimation of the Leave-Out Means type. It estimates the effect of the crash on the aggregate variables at the team level, discounting the individual contribution of each player. Although only the coefficients associated with key passes and shots (*XGChain* marginally significant) turn out to be statistically significant, with a drop in the total performance of the team of around 3.77% and 2.9%, respectively, in this case, it is extremely interesting to show the signs and magnitudes of the four groups in question. First, for groups of players with low productivity, I see that the relative effects are much smaller and close to zero compared to those with high productivity. Then, I see that the estimated sign is positive for the most productive players in productive teams in all the variables. This suggests that, in this type of team, the injuries of productive players are replaced mainly by players who are also effective, which means that, by discounting his contribution, the rest of the players have a positive effect due to the entry of a substitute player and/or the increase in the contribution of the rest of the players.

On the other hand, the impact on productive players in low-productivity teams is negative and significant in some cases. In this case, it gives an intuition of complementarity of this productive player over his teammates, i.e., discounting the decrease in the individual productivity. The rest of the players have worse performance since the affected player enhanced the performance of his teammates on the field.

Finally, Table 8 presents an estimate of both individuals' crossed elasticities, that is, how each player is affected by the absence due to injury of another, and aggregate elasticities,

Table 7: *Leave-Out* Means estimation at the team level.

	(1) Key passes	(2) XGChain	(3) NPXG	(4) Shots	(5) Expected Assistances	(6) Expected Goals	(7) XGBuildUp
(a) Low P. - Low T.	0.0792	0.0215	0.00679	0.0346	0.0142	0.0233	0.00162
	(0.116)	(0.0526)	(0.0129)	(0.117)	(0.0108)	(0.0160)	(0.0492)
(b) Low P. - High T.	-0.0271	-0.0992	-0.00595	-0.0515	-0.00606	0.00895	-0.0866
	(0.198)	(0.102)	(0.0400)	(0.295)	(0.0321)	(0.0418)	(0.0619)
(c) High P. - Low T.	-0.290**	-0.153*	-0.0296	-0.305**	-0.0283	-0.0222	-0.118
	(0.0955)	(0.0701)	(0.0170)	(0.0731)	(0.0145)	(0.0204)	(0.0564)
(d) High P. - High T.	0.310	0.153	0.0352	0.262	0.0376	0.0468	0.0824
	(0.231)	(0.0873)	(0.0320)	(0.329)	(0.0300)	(0.0274)	(0.0679)
Constant	8.620***	3.310***	1.137***	11.58***	0.878***	1.232***	1.965***
	(0.0735)	(0.0664)	(0.0140)	(0.0874)	(0.00836)	(0.0159)	(0.0535)
Observations	592,617	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.171	0.214	0.179	0.170	0.171	0.174	0.205
Mean at t-1	8.780	3.490	1.180	11.77	0.910	1.270	2.080
Mean (a)	7.920	2.740	1.020	10.84	0.760	1.100	1.530
Mean (b)	10.01	4.170	1.320	13.17	1.050	1.470	2.580
Mean (c)	7.680	2.890	1.020	10.43	0.800	1.080	1.690
Mean (d)	10.93	5.220	1.570	14.25	1.230	1.710	3.300

Note: Estimates include individual, match week-season fixed effects. Cluster standard errors at team level, game week-season level. (a) low productivity player - low productivity team. (b) low productivity player - high productivity team. (c) High productivity player - low productivity team. (d) high productivity player - high productivity team. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

whose interpretation is the aggregate effect on the team. I see how the values are smaller compared to groups of less productive players and can change significantly depending on the variable. However, when it comes to good players on unproductive teams, the robust positive sign reinforces the idea of complementarity between this class of players and their teammates. In contrast, for good players in productive teams, the negative sign reflects the substitution that may exist where most of the workforce is highly productive.

Table 8: Cross elasticities

	(1) Key passes	(2) XGChain	(3) NPXG	(4) Shots	(5) Expected Assistances	(6) Expected Goals	(7) XGBuildUp
				Panel A - Individual Elasticities			
(a) Low P. - Low T.	-0.0141	-0.0112	-0.0088	-0.0049	-0.0200	-0.0349	-0.0017
(b) Low P. - High T.	0.0031	0.0425	0.0088	0.0059	0.0106	-0.0108	0.0535
(c) High P. - Low T.	**0.0853**	**0.1050**	**0.0515**	**0.0546**	**0.0860**	0.0346	**0.1486**
(d) High P. - High T.	**-0.0701**	**-0.0790**	**-0.0564**	**-0.0502**	**-0.0780**	**-0.0740**	**-0.0652**
				Panel B - Aggregate Elasticities			
(a) Low P. - Low T.	-0.4125	-0.3080	-0.2978	-0.1436	**-0.7594**	**-0.9588**	-0.0372
(b) Low P. - High T.	0.2203	**1.4784**	0.2888	0.2861	0.5560	-0.3978	1.9727
(c) High P. - Low T.	**0.8978**	**1.3190**	**0.6563**	**0.6778**	**0.9826**	0.4150	**1.9313**
(d) High P. - High T.	**-0.8908**	**-1.0851**	**-0.7379**	**-0.6329**	**-0.9592**	**-0.9730**	**-0.9786**

Nota: En el panel A, las elasticidades cruzadas individuales se calculan como la división entre el efecto relativos de las estimación *Leave-Out Means* a nivel de equipo sobre el efecto relativo de la estimación individual, elasticidades mayores a 0.05 en valor absoluto en negrita. En el panel B, las elasticidades cruzadas agregadas se calculan como el coeficiente asociado al efecto de la lesión en la estimación *Leave-Out Means* sobre el coeficiente de la estimación individual, elasticidades mayores a 0.6 en valor absoluto en negrita.

6 Conclusions

This document provides new evidence on the effect of a decrease in the individual productivity of one of the productive factors on the total production in team contexts. Using exogenous events determined by unforeseen injuries to soccer players, a significant decrease in individual performance is identified in the periods after the shock, measured through different variables such as goals, expected assists, or modern measures of game creation. At the same time, no effect is found on the aggregate performance of the team. However, dividing the sample into teams and players according to their level of productivity, it is found that productive players on less productive teams have a negative relative effect of approximately 6%-8% on total shots, expected assists, anticipated goals, and generation of team play. When calculating the cross-elasticities for the four types of groups identified, I see how it is not entirely conclusive for the groups of low-productive players due to values close to zero and changes in signs depending on the analyzed variable. However, for players highly productive in low-productivity teams, a positive sign indicates the complementarity of this type of player. In contrast, the most productive players in highly effective teams show a negative sign that indicates possible substitutability between this player and his teammates, which by definition belongs to a team with a higher proportion of productive players.

6.1 Future Work

This article lays the groundwork for a research agenda on the individual contribution of factors of production to aggregate output in team contexts. Additional steps for the future are the download and use of defensive data, which allows us to have a comprehensive measure of player performance and to calculate cross-role interactions within the field of play, analyzing, for example, the effect on offensive productivity of the loss of defensive players.

Likewise, future work should focus on calculating confidence intervals for the elasticities presented in this report, either through a *Bootstrap-type* estimation or a 2-stage estimation that calculates the denominator and numerator of such elasticities together. Finally, it is in the plans to carry out robustness exercises on the inherent assumptions of the *event study*, in particular the assumption of homogeneity of the effect.

References

Abraham, S. & Sun, L. (2018), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *arXiv preprint arXiv:1804.05785* .

Arcidiacono, P., Kinsler, J. & Price, J. (2017), 'Productivity spillovers in team production: Evidence from professional basketball', *Journal of Labor Economics* **35**(1), 191–225.

Bonhomme, S. (2020), 'Teams: Heterogeneity, sorting and complementarity', *A preliminary draft of the work presented at the 2020 World Congress of the Econometric Society* .

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D. & Singleton, C. (2020), 'Experimental effects of an absent crowd on performances and refereeing decisions during covid-19', *Available at SSRN 3668183* .

Bundesliga (2018), 'xg stats explained: The science behind sportec solutions' expected goals model', https://www.bundesliga.com/en/bundesliga/news/expected-goals-xg-model-what-is-it-and-why-is-it-useful-sportec-solutions-3177.

Caruso, R., Di Domizio, M. & Rossignoli, D. (2017), 'Aggregate wages of players and performance in italian serie a', *Economia Politica* **34**(3), 515–531.

Devereux, K. (2018), Identifying the value of teamwork: Application to professional tennis, Technical report, Working Paper Series.

Dvorak, J., Junge, A., Graf-Baumann, T. & Peterson, L. (2004), 'Football is the most popular sport worldwide', *American journal of sports medicine* **32**(Suppl 1), 3S–4S.

Efe (2020), 'Las cinco grandes ligas europeas, cada vez más poderosas económicamente', https://www.mundodeportivo.com/futbol/internacional/20200117/472930369346/las-cinco-grandes-ligas-europeas-cada-vez-mas-poderosas-economicamente.html.

Ernst&Young-LLP (2019), Premier league: Economic and social impact, Technical report, Ernst Young LLP.

Freyaldenhoven, S., Hansen, C. & Shapiro, J. M. (2019), 'Pre-event trends in the panel event-study design', *American Economic Review* **109**(9), 3307–38.

Gauriot, R. & Page, L. (2019), 'Fooled by performance randomness: Overrewarding luck', *Review of Economics and Statistics* **101**(4), 658–666.

Gaviria, A. (2000), 'Is soccer dying? a time series approach', *Applied Economics Letters* **7**(4), 275–278.

Jacobson, L. S., LaLonde, R. J. & Sullivan, D. G. (1993), 'Earnings losses of displaced workers', *The American economic review* pp. 685–709.

Matano, F., Richardson, L. F., Pospisil, T., Eubanks, C. & Qin, J. (2018), 'Augmenting adjusted plus-minus in soccer with fifa ratings', *arXiv preprint arXiv:1810.08032* .

Neale, W. C. (1964), 'The peculiar economics of professional sports', *The quarterly journal of economics* **78**(1), 1–14.

Rosen, S. & Sanderson, A. (2001), 'Labour markets in professional sports', *The economic journal* **111**(469), F47–F68.

Roth, J. (2018), Pre-test with caution: Event-study estimates after testing for parallel trends, Technical report, Working Paper.

Rottenberg, S. (1956), 'The baseball players' labor market', *Journal of political economy* **64**(3), 242–258.

Scully, G. W. (1995), *The market structure of sports*, University of Chicago Press.

Sloane, P. J. (1971), 'Scottish journal of political economy: the economics of professional football: the football club as a utility maximiser', *Scottish journal of political economy* **18**(2), 121–146.

Stubbe, J. H., van Beijsterveldt, A.-M. M., van der Knaap, S., Stege, J., Verhagen, E. A., Van Mechelen, W. & Backx, F. J. (2015), 'Injuries in professional male soccer players in the netherlands: a prospective cohort study', *Journal of athletic training* **50**(2), 211–216.

Zuke, W. A., Agarwalla, A., Go, B., Griffin, J. W., Cole, B. J., Verma, N. N., Bach, B. R. & Forsythe, B. (2018), 'The lack of standardized outcome measures following lower extremity injury in elite soccer: a systematic review', *Knee surgery, sports traumatology, arthroscopy* **26**(10), 3109–3117.

7 Appendix A

Table 9: Effect of injury on individual variables

	(1) Assisances	(2) Yellow card	(3) Red card	(4) NPG	(5) NPGX	(6) XGchain
Pre	-0.00194	0.0217	0.00174	0.0140	0.00719	0.00986
	(0.00499)	(0.0119)	(0.000992)	(0.00838)	(0.00608)	(0.00781)
Post	-0.0303***	-0.0474**	0.00102*	-0.0279**	-0.0334***	-0.0965***
	(0.00487)	(0.0128)	(0.000384)	(0.0100)	(0.00650)	(0.0120)
Constant	0.0521***	0.0819***	0.000583	0.0496***	0.0550***	0.163***
	(0.00450)	(0.0114)	(0.000587)	(0.00803)	(0.00520)	(0.00939)
Observations	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.065	0.062	0.010	0.108	0.215	0.222
Mean at t-1	0.0600	0.100	0	0.0600	0.0600	0.190

Note: Estimates include individual, match-week-season fixed effects. Cluster standard errors at the individual, game week-season level. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

Table 10: Effect of injury on team variables

	(1) Tie	(2) Victory	(3) Defeat	(4) Goal Difference	(5) Scored Goals	(6) Conceded Goals
Pre	-0.00670	0.0254	-0.0186	0.0468	0.00298	-0.0438
	(0.0195)	(0.0134)	(0.0174)	(0.0524)	(0.0335)	(0.0387)
Post	-0.00770	0.0141	-0.00644	-0.0175	-0.0496	-0.0321
	(0.0224)	(0.0192)	(0.0270)	(0.0586)	(0.0327)	(0.0494)
Constant	0.253***	0.369***	0.378***	0.0281	1.403***	1.375***
	(0.0221)	(0.0197)	(0.0231)	(0.0620)	(0.0423)	(0.0422)
Observations	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.022	0.086	0.071	0.118	0.110	0.060
Mean at t-1	0.250	0.390	0.370	0.100	1.470	1.370

Note: Estimates include individual, matchweek-season fixed effects. Cluster standard errors at team, game week-season level. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

Table 11: Heterogeneous effects of injury on individual variables

	(1) Assisances	(2) Yellow card	(3) Red card	(4) NPG	(5) NPGX	(6) XGchain
(a) Low P. - Low T.	-0.0249***	-0.0513**	0.000537	-0.0151	-0.0228**	-0.0698***
	(0.00475)	(0.0160)	(0.000748)	(0.00805)	(0.00504)	(0.00858)
(b) Low P. - High T.	-0.0173*	-0.0213	0.00118	-0.0174	-0.0206*	-0.0671***
	(0.00699)	(0.0189)	(0.00112)	(0.0100)	(0.00889)	(0.0144)
(c) High P. - Low T.	-0.0381***	-0.0531***	0.00146	-0.0390**	-0.0451***	-0.116***
	(0.00686)	(0.00750)	(0.000826)	(0.0117)	(0.00739)	(0.0124)
(d) J. Alta - E. Alta	-0.0392***	-0.0483**	0.00134	-0.0458**	-0.0477***	-0.141***
	(0.00691)	(0.0138)	(0.00122)	(0.0152)	(0.0101)	(0.0180)
						(0)
Constant	0.0506***	0.0812***	0.000577	0.0484***	0.0540***	0.160***
	(0.00444)	(0.0111)	(0.000601)	(0.00830)	(0.00535)	(0.00955)
Observations	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.065	0.062	0.010	0.108	0.216	0.223
Mean at t-1	0.0600	0.100	0	0.0600	0.0600	0.190
Mean (a)	0.0200	0.100	0	0.0200	0.0300	0.100
Mean (b)	0.0200	0.0700	0.0100	0.0500	0.0400	0.120
Mean (c)	0.0800	0.100	0	0.0600	0.0800	0.230
Mean (d)	0.150	0.100	0	0.140	0.120	0.380

Note: Estimates include individual, game week-season fixed effects. Cluster standard errors at the individual, game week-season level. (a) low productivity player - low productivity team. (b) low productivity player - high productivity team. (c) High productivity player - low productivity team. (d) high productivity player - high productivity team. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.

Table 12: Heterogeneous effects of injury on team-level variables

	(1) Empate	(2) Victoria	(3) Derrota	(4) Gol Diferencia	(5) Goles Anotados	(6) Goles Concebidos
(a) Low P. - Low T.	-0.00582	0.0174	-0.0116	0.0247	-0.0268	-0.0515
	(0.0257)	(0.0242)	(0.0249)	(0.0596)	(0.0435)	(0.0378)
(b) Low P. - High T.	-0.0310	0.0276	0.00338	-0.0137	-0.0308	-0.0171
	(0.0216)	(0.0339)	(0.0387)	(0.117)	(0.0651)	(0.0809)
(c) High P. - Low T.	-0.00654	-0.00119	0.00773	-0.0871	-0.0824	0.00476
	(0.0177)	(0.0239)	(0.0292)	(0.0807)	(0.0412)	(0.0557)
(d) High P. - High T.	0.000297	0.0176	-0.0179	-0.0223	-0.0668	-0.0445
	(0.0342)	(0.00830)	(0.0310)	(0.0491)	(0.0380)	(0.0652)
Constant	0.253***	0.369***	0.378***	0.0258	1.401***	1.375***
	(0.0224)	(0.0196)	(0.0236)	(0.0617)	(0.0400)	(0.0424)
Observations	592,617	592,617	592,617	592,617	592,617	592,617
R-Squared	0.022	0.086	0.071	0.119	0.110	0.060
Mean at t-1	0.250	0.390	0.370	0.100	1.470	1.370
Mean (a)	0.250	0.320	0.430	-0.290	1.230	1.520
Mean (b)	0.310	0.430	0.270	0.490	1.670	1.180
Mean (c)	0.280	0.270	0.450	-0.250	1.270	1.510
Mean (d)	0.180	0.610	0.210	1	2.030	1.030

Note: Estimates include individual, match week-season fixed effects. Cluster standard errors at team level, game week-season level. (a) low productivity player - low productivity team. (b) low productivity player - high productivity team. (c) High productivity player - low productivity team. (d) high productivity player - high productivity team. *, **, and *** indicate significance level at 10%, 5%, and 1%, respectively.