“There is no ‘I’ in Team”
Skill Distribution and Team Performance: Evidence from the German Bundesliga

Mario Jametti
Università della Svizzera italiana (USI), Switzerland and CESifo

Eleonora Lanzio
Università della Svizzera italiana (USI), Switzerland

Edoardo Slerca
Università della Svizzera italiana (USI), Switzerland

Incomplete draft - do not cite nor distribute
February 15th, 2019

1 Introduction

Economic performance relies in almost all cases on a collaborative effort. Societies decide on the allocation of talent to foster economic growth, firm management seeks the optimal organizational structure to maximize profits and work teams assign tasks to improve production. An important question in this regard is to what extent the composition of the collaborative effort affects economic performance. Hence, abstracting from the multi-dimensional aspects of team composition, should organizations strive for homogeneous or dispersed groups? In other words, should societies aim and invest in a fairly uniform talent pool to foster economic growth? Or should they, instead, strive to diversify available competencies or skills. Similarly, should firms and organizations combine similar skills in work groups? Or does a diverse portfolio of skills increase performance.

In this paper we aim to shed some light on these questions entering the realm of sports. For this, we analyse the impact of the skill distribution among players on the overall performance of a team. Using data from the German Bundesliga we explore whether, *ceteris paribus*, a more equal distribution of skills within a team leads to a better performance, and whether this effect might be non-linear. More specifically, data availability for sports allows us to use a one-dimensional proxy for skills, namely a player's market value. This proxy serves us to construct skill distribution measures by team in the league and to analyze its effect on performance, controlling for many other factors, including the overall market value of a team.

Many of the early contributions concentrated on the relationship between wage dispersion and economic performance. The theoretical effect there is ambiguous. On the one hand, *Equity Theory* formulated in the seminal papers by Akerlof and Yellen (1988), Akerlof and Yellen (1990) and Levine (1991) claim that relative wage equality is beneficial for people working in teams. On the other hand, *Tournament Theory* proposed by Lazear and Rosen (1981) and Rosen (1986), sustains that workers in firms behave like players competing in a tournament whose prize is a higher salary, and that therefore more disparity leads to better performance. Much less is known on what the effect of skill composition in a team is.

In order to structure our empirical estimation, we build a simple model of group organization and its impact on (economic) outcome. The model allows for both complementarities and substitutabilities of skill composition, to guide the empirical implementation. Theory suggests that the effect of skill distribution on performance is likely to be non-linear.

The dataset at hand consists of season and match level information of the *Bundesliga* for the period 2003-2004 to 2015-2016, combined with time-varying individual valuations of all players active in the league during this period. The latter information is used to construct team-level variables of market value distribution, namely a *Gini Heterogeneity Index (GHI)*, as well as the overall market valuation of the team.

We then regress performance measures (e.g. points), both at the seasonal and individual game level, on overall market valuation and our measure of skill distribution. In the match specification we are able to include other relevant control variables such as home-game, etc., while the panel format of our data allows us to use an array of fixed effects, including club and season effects, in all specifications.

As theory suggest, the effect of skill distribution might be non-linear. As such, we allow for non-linear parametric as well as semi- and non-parametric effects. Further, to control for the potential reverse causality of performance on team valuation and distribution, we apply a shift-share instrumental variable approach.

2 Literature review

Analogously to the theoretical literature, the previous work on the relationship between wage dispersion and performance in sports came to divergent conclusions: following the strand of Equity Theorem, Depken II (2000) found that in North American baseball teams, greater wage disparity has a negative effect on team performance. Frick et al. (2003), in a study comparing seasonal performances in North American major leagues showed that wage disparity could have both a positive (in hockey and basketball) or a negative (in American football and baseball) impact on performance. The authors suggest that the different sign of the impact depends on the size of the squad on the field. Katayama and Nuch (2011), using NBA data, argue that there is no causal effect of wage disparity on performance. Conversely, using data from the NBA as well, Simmons and Berri (2011) find that *"expected pay dispersion has a positive impact effect on team and individual performance"*, linking their result with Tournament Theory. Finally, using estimates from the web page of a German sports magazine as a proxy for salary, Franck and Nüesch (2011) find that there is a non linear relationship between performance of football players and wage dispersion. Moreover they highlight how it is not clear *a priori* whether that relationship is U-shaped or hump-shaped.

Even though wage has been used extensively in the sports literature, some authors suggest that it might give a distorted measure of skills: factors such as media coverage (Lehmann and Schulze, 2008), *"winner-take-it-all effect"*(Garcia-del Barrio and Pujol, 2004) and *"superstar effect"* (Lucifora and Simmons, 2003) make salaries grow more than linearly in skills. Hence, to address this issue we use players' market values, instead. We would argue that this is a more reliable proxy for of skills. One important aspect is that it includes both present and future expected performances, accounting for a risk factor. The risk factor is an endemic component given that players might be subject to injuries, or might be unable to fully express in the future their expected potential.

3 Brief overview for the (rare) non football lovers

In the German *Bundesliga* is the premier league of professional football in Germany. It consists of 18 teams competing each season, which lasts from August to May with a winter break from December to January. Teams play against each other twice in the season, once at home and once away. A match lasts two halves of 45 minutes each. The referee has the discretion to extend playing time to compensate for time lost for any reason during the match, i.e. the clock is not stopped during interruptions. The objective of the game is to score goals to obtain points for the championship. Thus, at the end of the match a club is assigned zero points for a loss, one point for a tie and three points for a win. The club that at the end of the season gets more points is awarded the title of league champion (*Deutsche Meisterschale*), with special provisions in case of a tie of points.

The ranking in the league is relevant for a number of reasons. Most importantly and contrary to most North American sports leagues, the *Bundesliga* follows the traditional (European) scheme of promotion and relegation. The last two clubs in the ranking are directly demoted to the lower league (*2. Bundesliga*), while the third-last ranked club plays a two-game tiebreaker with the third classified in the lower league. Further, ranking determines participation in European club-football, namely the Champions and Europa Leagues organised by UEFA, in the next season. Currently, the four first teams compete for the UEFA Champions League, while the fifth and sixth ranked participate in the lesser ranked Europa League.[1]

Similar to many other sports leagues in the world, there exist also for the German Bundesliga a plethora of individual player statistics which are easily accessible. Most importantly for our analysis, for each player a periodically updated market value is available online, published by the website *transfermarkt.de*. For each active player, the market value is update at least twice

[1]For completeness, the German Football Federation organizes a parallel cup competition, where teams from different leagues play each other and only the winner advances to the next stage. The winner of the national cup competition is also qualified for the Europa League. Winning both the championship and the cup in one season is called a "double".

per season. It should be noted that we are agnostic on the methodology applied to calculate the market values, as such we take them as face value for our proxy of skills. However, this data is often used as the reference point used by the European press and by insiders when assessing or comparing both players and teams. For example, for a particular match relative strength of teams is often assessed by the overall market values of players in a team.

4 A simple model

In order to structure our empirical estimation, we build a simple model of group organization and its impact on (economic) outcome. The model allows for both complementarities and substitutabilities of skill composition, to guide the empirical implementation. Theory suggests that the effect of skill distribution on performance is likely to be non-linear.

The model uses a production function approach which includes different types of labour skills. Both the level of skills available, as well as its combination, affect output or performance, the latter in a non-linear way. Hence, the theoretical model boils down to the following equation:

$$\begin{aligned} performance_c = \beta_0 + \beta_1 SkillLevel_c + \\ \boldsymbol{f(SkillDistribution_c)} + + \boldsymbol{X'\beta} + \epsilon_c. \end{aligned} \tag{1}$$

The performance of team c is a function of the overall level of skills, its distribution and potential other factors. Given that we use market values as a proxy to construct the level and distribution of skills, it will be important to consider the fact that current performance of a team likely might influence market valuation, hence creating a reverse causality problem, to which we return below.

TBC

5 Data

The basis for our panel-dataset comes from the combination of two sources. Firstly, we collected from the official German Football Federation website *dfb.de* match level information on 13 seasons (from season 2003-2004 to 2015-2016) of the German Bundesliga. As mentioned before, the league consists of 18 teams per season for a total of 34 teams in the period. Information contained in this source includes the outcome of the game, goals scored, location information (home vs. guest) as well as match specific data, such as information on scorers, red cards, trainers and referees.

Secondly, from the website *transfermarkt.de*, we gathered market value for players that set foot on the field at least once in a *Bundesliga* game during the seasons we analyse. We also obtained information on the precise position of each player, which we aggregate in the four main positions for football, namely: goalie, defender, centerfielder and striker. Lastly we complemented the dataset with information on historical rivalries and participation in the European wide Champions and Europa Leagues' international competitions.

The structure of the dataset is the following. For each match we have two observations: one for the home and one for the guest club. For each club and each player in the starting roster (eleven of them) we have a variable for player id, role and market value (at the time of the game). We further have the same information for each substituted player, as well as the time he entered (and whom he substituted).

The information on the matches allows us to calculate the rank and accumulated points at the time of each match, as well as the difference in rank and points between the opponents. Finally, we have detailed information on events on the field (goals, red cards, yellow cards) and the name of the trainers and the name of the referee.

We use the structure of the dataset for our two main time-frame specifications: for the *match specification* we use all observations in the dataset, while for the *season specification* we consider only the last round in the championship.

5.1 Dependent variables

Our dependent variable is the number of points that each team gets after each season (for the season specification) or after each match (for the match specification). We opted for the number of points as a measure of performance rather than the number of goals scored since the number of points is the one mostly relevant for the final ranking and therefore for the determination of the national champion.

5.2 Independent variables

From a first elaboration of the data we were able to construct the dependent variables used in our empirical specifications. These are structured in three categories, which we in turn describe: *i*) the main dependent variables, i.e. market value indicators and dispersion measures;

ii) instruments to deal with the potential reverse causality of performance on players' market values and on the dispersion index; and iii) additional controls to enhance the precision of our analysis.

5.2.1 Main dependent variables

Team Market Value

From the starting 11 players on the field, from the timing of substitutions, from the double yellow cards and from the red cards, we constructed a variable (mof) stating the exact number of minutes that each player spent on the field.[2] For each match, moreover, we have information of the exact date of the match. This allowed us to attribute to each player the current market value, allowing for adjustments in the quotation of each individual throughout the season. The market value of a team for a specific match is then the sum of the values of each player in the field, weighted by mof as a percentage of the total playing time, i.e. 90 minutes:

$$MarketValue_{c,m} = \sum_{i=1}^{n} \left(\frac{mof_i}{90} * value_i \right),$$

where c denotes the club, m the match and n is a number between 11 and 14 depending on the number of substitutions that the trainer decided on during the match.

For the seasonal specification we use the value of all players that entered at least one game during the championship.[3] Further, in case of a change in a player's market value within the season, we use its minimum to be more conservative.

$$MarketValue_{c,s} = \sum_{i=1}^{k} (value_i),$$

where s indicates the season and k is the number of player on the roster. Finally, in order to account for possible decreasing marginal returns to value we use the log of $MarketValue$ in all specifications.

Dispersion measures

To measure skill distributions we opted for the Gini Heterogeneity Index GHI_x. The GHI is defined as 1 minus the squared sum of shares of the the market values of each player, i.e. 1 minus the Hirschmann-Herfindahl Index (HHI) of the players' value distribution.

[2] We capped mof at 90 minutes.

[3] Our definition might be slightly different from squads presented by clubs, as we do not consider players that train with the squad but never enter a game. A good example for this might be the second reserve goalie.

The Gini Heterogeneity Index per season ($GHI_{c,s}$) captures the skill distribution of the roster. In this case we put an equal weight on the market value of each player no matter for how long they actually played.

$$GHI_{c,s} = 1 - \sum_{i=1}^{k} \left(\frac{value_i}{\sum_{i=1}^{k} value_i} \right)^2 .$$

where k is the again number of players composing roster.

The Gini Heterogeneity Index per club and per match ($GHI_{c,m}$) captures the skill distribution of the team put on the field and is computed in the following way:

$$GHI_{c,m} = 1 - \sum_{i=1}^{n} \left[\frac{\frac{mof_i}{90} * value_i}{\sum_{i=1}^{n} \left(\frac{mof_i}{90} * value_i \right)} \right]^2 .$$

The Gini Heterogeneity Indexes can therefore assume values ranging from 0 to 1. The higher the GHI the **less** dispersion there is among players' market values. A team with a GHI close to 1 would be a perfectly egalitarian one.

5.2.2 Instruments

As mentioned before, our regressions might be plagued by reverse causality, i.e. the effect of team performance on individual players' market value and its distribution in the team, rendering OLS estimates biased. The problem is more acute for the seasonal specification, while less so for the match specification, as market values are only intermittently adjusted.

The instrument we constructed is inspired by the large literature of shift-share instruments *a la Card*.

For the instrument we considered that each team is composed of four different positions, namely: goalkeepers, defenders, centerfielders and strikers. We then used the cross-sectional variation in market values for the first season in our dataset and projected this distribution by the league-wide growth rate of values for each position.

Specifically, we tackled the problem of an instrument for the market values as follows.

From the panel dataset with the detailed game level information we constructed a second database insulating individual player data. This database contains only one market value per player per club per season. To be more conservative we used the minimum value registered for each player playing for that particular club in that particular season. [4].

[4] A player that is transferred mid-season can, in principle, have more than one value during the season, but only one per club per season

We then created a variable $average\_value_{s=1,p}$ that contained the average market value by position p for the first season $s = 1$. Afterwards we created a variable $average\_value_{s,p}$ containing the average market value by position p for all seasons s.

To avoid problems of self instrumentation we excluded the values of team c when computing both $average\_value_{s,p}$ and $average\_value_{s=1,p}$. In this way we were able to compute for each club (by dividing $average\_value_{s,p}$ by $average\_value_{s=1,p}$) a league-wide value growth rate by position with respect to the first season that excluded the club itself.

Finally we created a variable $average\_value_{s=1,c,p}$ containing the average value by position p by club c for the first season $s = 1$.

To each player i in each season we assigned a value $shift\_av\_value_{s,c,i}$ equal to the average value of his position in the first season, multiplied by the growth rate of the value of his position in the season he is currently playing with respect to the initial season:

$$Shift\_av\_value_{s,c,i} = average\_value_{s=1,c,p} * \frac{average\_value_{s,p}}{average\_value_{s=1,p}}.$$

Summing them over the clubs, we were able to identify an instrumental variable for the market value of the club roster in each season $shift\_value_{s,c}$:

$$Shift\_Market\_Value_{s,c} = \sum_{i=1}^{k} shift\_av\_value_{s,c,i},$$

where k is, as before, the number of players composing the roster.

Since we are dealing with an unbalanced panel with clubs entering and exiting the competition in each season, the construction of the instrument faced an additional challenge. This comes from the fact that, for clubs that entered the *Bundesliga* in subsequent seasons (and that therefore were not present in the first season of our dataset) we are not able to construct $average\_value_{s=1,c,p}$. To solve this issue we computed the average value by position for the last three clubs in the *Bundesliga* ranking in the first season, and we assigned it as $average\_value_{s=1,c,p}$ for the clubs that entered the league later.

The logic behind is that the closest value for someone that is promoted to the *Bundesliga*, is the one of the teams that got relegated (up to the last three in the league). By doing so we were able to compute the instruments for average market value and for total roster value also for the clubs that were not present in season 2003 - 2004.

Once we constructed the IV for the market value, all we have to do was to use it to construct the IV for the $GHI_{c,s}$ adapting the formula previously used for the construction of the initial dispersion measures:

$$Shift\_GHI_{c,s} = 1 - \sum_{i=1}^{k} \left(\frac{shift\_av\_value_{s,c,i}}{\sum_{i=1}^{k} shift\_av\_value_{s,c,i}} \right)^2 .$$

The instrumental variable for the Gini Heterogeneity Index IV_GHI for club c in season s is equal to 1$-$ the shifted average market value of player i in club c and season s over the shifted total value of the roaster of his club in the same season, squared.

5.2.3 Additional controls

For each club and each match we had individual information on goals (we knew who scored and when during the match). From this information we computed a variable indicating the number of goals scored by each club during the game, with which we were able to compute the total number of goals scored by each club in previous matches, the goal difference (scored/received) of each club before the game, the points of each club before each game and subsequently its ranking before each match. We constructed also a variable indicating the points of the previous match for each club that tells us if the club comes from a win or a loss (which could be relevant in terms of morale), and a variable indicating the points difference with the opponent before the match, that indicates the disparity in past performances among clubs and therefore can represent an expectation of future performance during the match.

As stated before, for each match we have information of the exact date of the match. This was relevant for the construction of dummies about the current involvements of clubs in other international competitions: a dummy indicates that the club is currently playing in Champions League, while another dummy states if the club is involved in the Europa League at the moment of the match).

Another element that could influence performance is whether there is something important at stake: we created dummies for the teams that, given their current position in the ranking, can aim to the Europa League and the Champions League and are therefore in the race for first place in the Championship. Analogously we crated a dummy for teams that are risking relegation.

In addition to these controls we have information on trainers and referees, which are used in the match specification as fixed effects alongside club and season ones (the latter are used in the seasonal specification as well).

Summary statistics are presented in Table (1) and (2). Note that the average of $MarketValue$ is significantly higher for the season as opposed to the match specification. Recall that this is due to the inclusion of all players on the roster throughout the season, while for each match only a maximum of 14 players (starting squad plus substitutions) is considered.

6 Empirical implementation

6.1 Baseline model

We run specifications both at the season and match level. For the seasonal specification the basic regression equation is:

$$\begin{gathered} points_{c,s} = \beta_0 + \beta_1 MarketValue_{c,s} + \\ \boldsymbol{f(GHI_{c,s})} + \kappa_c + \gamma_s + \epsilon_{c,s}, \end{gathered} \tag{2}$$

Where points that a club c accumulates during a season s are influenced by the total market value of the roaster, a function of the dispersion index (GHI) at the seasonal level, club fixed effects κ and a year γ fixed effects.

The match specification, instead, allows to account for a richer set of controls:

$$\begin{gathered} points_{c,m} = \beta_0 + \beta_1 MarketValue_{c,m} + \boldsymbol{f(GHI_{c,m})} \\ + \boldsymbol{X'\beta} + \kappa_c + \gamma_s + \tau_{c,m} + \rho_m + \epsilon_{c,s} \end{gathered} \tag{3}$$

In addition to the club and year fixed effects from before, we include trainer τ and referee ρ effects. The control variables are: the points that the team obtained in the previous match, the points difference before the beginning of the match, a home game dummy, dummies that specify whether the team is currently risking relegation or whether they are competing to enter Europa or Champions League the following season, dummies whether the team is currently involved in Champions League or Europa League at the time of the match, and dummies on the teams on the field being involved in historical rivalries.

6.2 Specifications

Besides the season and match specifications, we run all estimations also on a reduced dataset including only the teams that throughout our period of analysis were present in the *Bundesliga*. There are a total of nine such teams. We do this because our shift-share instrument is more straightforward for these teams.

Two items remain to be discussed. The functional form of $\boldsymbol{f(GHI_{c,m})}$ and the IV specification.

Regarding the GHI we start with a parametric implementation, going from linear to include up to the fourth power. In order to allow for more flexibility on the functional form of the effect of the distribution index we also perform semi-parametric panel regressions, with a non-parametric estimation of GHI.

Finally, currently we conducted the instrumental variable analysis solely at the seasonal level, but we plan to do it at a match level as well in the near future. As mentioned before, the potentially endogenous variables are the market value and the dispersion index.

7 Results

7.1 Seasonal specifications

7.1.1 Functional Form

Table (3) presents the season level OLS-regressions on the full sample (with a total of 34 teams). All specifications include team and season fixed effects. Columns (1) and (2) present simple regression with a linear GHI. By itself, as expected, a higher team-level market value increases performance, measured by the number of points obtained in the championship. As such, money buys success. The effect is sizeable since the semi-elasticity, columns (3) to (6), hoovers around 11.5. This implies that a 1 percent increase in $MarketValue$ (around 8 million for the average team) increases the points total by more than eleven points (evaluated at the mean this implies almost 25 percent more points).

- can we say how many ranks you gain?

Similarly, in isolation a more egalitarian skill distribution (a higher GHI) reduces team performance. Including both variables linearly, in column (3), confirms these results. The effect of GHI is an order of magnitude smaller than the one of $MarketValue$, i.e. a one standard deviation increase in GHI implies around three less championship points ($-59.57 * 0.05 \approx 3$).

However, as suggested by theory, the effect of the skill distribution proxy does indeed seem to be non-linear, presented in columns (4) to (6) including the various powers of GHI. In the quadratic specification, column (4), the main effect of GHI is positive while the coefficient on GHI^2 is negative, but smaller in absolute value. All coefficients are highly statistically significant. Interestingly, since GHI varies between 0 and 1, the marginal effect of GHI turns actually positive ($73.6 - 280.3 * GHI$) in column (4).

The third order specification, column (5) does not seem to fit the data very well, as none of the coefficients on the GHI are significant. Matters improve again with a quartic specification, where all four coefficients are highly statistically significant. Note that for all values of GHI in the sample, the overall effect is negative. Further, when plotting the marginal effect of the GHI on performance we observe a reverse U-shaped relationship. Thus, there is a local maximum of optimal skill heterogeneity around $GHI = 0.77$, roughly four percentage points below the average in the data. Hence, controlling for the level of skills, teams seem to be too egalitarian.

Figure 1 nicely illustrates this result plotting the marginal effect of GHI. The squares in the figure represent the average distribution (GHI) of the first three teams in the championship, while the dots are for the last three teams. We can see that the top three teams are close to the local maximum, and the bottom three are on average farther away.

We next turn to a semi-parametric estimation of our seasonal specification, allowing for even more flexibility for GHI.[5] Table (4) presents the output from the parametric part of the regression, while Figures (2 and (3 plot the corresponding effect of GHI, using a local polynomial and B-spline smooth, respectively. A couple of interesting aspects of these estimates. First, similar to the higher-order polynomial specification, the effect of skill distribution is flattened of the observed range of GHI, as compared to fitting a line. Second, the effect of GHI is positive for values up to about 0.83, where it turns negative.

A similar picture emerges from a raw kernel in the points-GHI scatterplot 4.

As a second robustness check Table (5), we performed a FE regression on the value of a team's roaster, computed the residuals and – subsequently – estimated non-parametrically the effect of the Gini index on these residuals. The effect of the value of a team's roaster is here positive and statistically significant, while the non-parametric effect of GHI on the residuals is negative and linearly so, Figure 5.

7.1.2 Instrumental Variables

The results of our instrumental variables regressions are presented in Table (6), with the corresponding first-stage results in Table (7). We only estimated the IV regression on the linear specification for GHI, since our instrument does not have enough power for a non-linear specification.

Qualitatively, our earlier results are confirmed, with slightly larger coefficients, see column (3). A higher $MarketValue$ has a positive effect on the number of Championship points, while

[5]We use the Stata *semipar* command, estimating GHI non-parametrically.

a more egalitarian skill distribution has a negative effect.

Our shift-share instruments perform surprisingly well. From the first-stage regressions, columns (3) and (4) of Table (7), we see that our instruments pass an ad-hoc weak instrument test, with F-statistics above 10.

Since, we used additional assumptions when generating the shift-share instrument for teams that were not always playing in the *Bundesliga*, we also run our IV-regression for the nine team that played at the highest level throughout our period. Results are presented in Table (8), with the first-stage results in Table (9).

Column (1) presents the OLS-results of the reduced sample, while columns (2) to (4) are the respective IV ones. Overall, the coefficient estimates are very similar and confirm our results. Hence, our IV-strategy does not seem to be driven by our assumption on promoted or relegated teams.

7.2 Match specifications

TBC

8 Discussion of Results

TBC

9 Conclusions and next steps

Our preliminary result up to now do suggest that there is indeed an optimal level skill heterogeneity, but the quadratic functional form should be relaxed.

We plan on conducting the semi parametric and IV analyses at a match level as well.

References

Akerlof, G. A. and Yellen, J. L. (1988). Fairness and unemployment. *The American Economic Review*, 78(2):44–49.

Akerlof, G. A. and Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics*, 105(2):255–283.

Depken II, C. A. (2000). Wage disparity and team productivity: evidence from major league baseball. *Economics letters*, 67(1):87–92.

Franck, E. and Nüesch, S. (2011). The effect of wage dispersion on team outcome and the way team outcome is produced. *Applied Economics*, 43(23):3037–3049.

Frick, B., Prinz, J., and Winkelmann, K. (2003). Pay inequalities and team performance: Empirical evidence from the north american major leagues. *International Journal of Manpower*, 24(4):472–488.

Garcia-del Barrio, P. and Pujol, F. (2004). Pay and performance in the spanish soccer league: who gets the expected monopsony rents. Technical report, Working Paper.

Katayama, H. and Nuch, H. (2011). A game-level analysis of salary dispersion and team performance in the national basketball association. *Applied Economics*, 43(10):1193–1207.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy*, 89(5):841–864.

Lehmann, E. E. and 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(1):59–70.

Levine, D. I. (1991). Cohesiveness, productivity, and wage dispersion. *Journal of Economic Behavior & Organization*, 15(2):237–255.

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

Rosen, S. (1986). Prizes and incentives in elimination tournaments. *The American Economic Review*, 76(4):701–715.

Simmons, R. and Berri, D. J. (2011). Mixing the princes and the paupers: Pay and performance in the national basketball association. *Labour Economics*, 18(3):381–388.

Tables and figures

Table 1: Season level - Summary Statistics

Season Variables	Obs	Mean	Std. Dev.	Min	Max
Points	243	46.782	13.965	18	91
Market Value	243	81.232	69.177	14.25	449.5
GHI	243	.819	.0500	.520	.9052
IV Market Value	243	80.033	58.950	16.585	317.304
IV GHI	243	.958	.005	.9295	.969

Table 2: Match level - Summary Statistics

Match Variables	Obs	Mean	Std. Dev.	Min	Max
Points	7,956	1.376	1.319	0	3
Market Value	7,956	51.938	45.854	4.751	347.3
GHI	7,956	.878	.0254	.634	.919
Points Previous match	7,722	1.375	1.317	0	3
Poins Difference With Opponent	7,956	0	11.985	-60	60
Guest	7,956	.500	.500	0	1
Relegation Risk	7,956	.384	.4865	0	1
Current Champions League	7,956	.116	.320	0	1
Current Europa League	7,956	.119	.324	0	1
Future Cups	7,956	.549	.498	0	1
Regional Rivalry	7,956	.043	.203	0	1
Interregional R ivalry	7,956	.026	.160	0	1
Local Rivalry	7,956	.001	.032	0	1

Table 3: Season level OLS - Including all observations

dependent variable: Points at the end of the season	(1)	(2)	(3)	(4)	(5)	(6)
Market Value	13.57***		11.51***	11.50***	11.42***	11.43***
	(2.747)		(2.709)	(2.534)	(2.507)	(2.443)
GHI		-77.15***	-59.57***	373.6***	-1,991	-47,760***
		(16.93)	(17.38)	(123.2)	(1,546)	(11,134)
GHI^2				-280.3***	2,992	98,874***
				(80.08)	(2,154)	(23,494)
GHI^3					-1,487	-89,689***
					(982.4)	(21,779)
GHI^4						30,106***
						(7,493)
Constant	-98.32***	109.0***	-28.83	-194.5***	366.5	8,445***
	(28.65)	(14.80)	(30.53)	(60.27)	(371.1)	(1,954)
Team FE	YES	YES	YES	YES	YES	YES
Season FE	YES	YES	YES	YES	YES	YES
Observations	234	234	234	234	234	234
Number of teams	34	34	34	34	34	34
R-squared	0.152	0.113	0.212	0.234	0.241	0.266

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 4: Seasonal level Semi Parametric estimation - Including all observations

dependent variable: Points at the end of the season	(1)	(2)
Market Value	-5.263	-4.747
	(4.312)	(4.252)
Team FE	YES	YES
Season FE	YES	YES
Observations	182	182
R-squared	0.137	0.135

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 5: Seasonal level 2-step procedure - Including all observations

dependent variable:	(1)
Points at the end of the season	
Market Value	16.21***
	(1.587)
Constant	-127.3***
	(17.02)
Team FE	YES
Season FE	YES
Observations	230
Number of teams	34

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 6: Season level IV - Including all observations

dependent variable:	(1)	(2)	(3)
Points at the end of the season			
Market Value	93.77**		18.01**
	(38.49)		(7.451)
GHI		-208.2***	-223.1***
		(31.55)	(38.81)
Constant	-961.3**	213.2***	31.23
	(413.3)	(25.89)	(99.46)
Team FE	YES	YES	YES
Season FE	YES	YES	YES
Observations	234	234	234
Number of teams	34	34	34

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 7: Season level IV - First stage - Including all observations

	(1)	(2)	(3)	(4)
	Market Value	GHI	Market Value	GHI
Shift Market Value	-0.508**		-2.074***	0.105***
	(0.248)		(0.490)	(0.0346)
Shift GHI		4.661***	51.72***	2.247**
		(0.618)	(11.62)	(1.072)
Constant	12.73***	-3.671***	-30.77***	-1.763*
	(0.990)	(0.593)	(9.487)	(0.917)
Team FE	YES	YES	YES	YES
Season FE	YES	YES	YES	YES
Observations	234	234	234	234
Number of teams	34	34	34	34
R-squared	0.420	0.339	0.531	0.360
F-statistics	4.2025	56.8516	10.17	37.47

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 8: Season level IV - Including clubs always present

dependent variable:	(OLS)	(1)	(2)	(3)
Points at the end of the season				
Market Value	15.93***	60.78***		25.05***
	(3.118)	(17.11)		(8.821)
GHI	-83.78***		-179.5***	-196.9***
	(21.95)		(56.77)	(49.53)
Constant	-56.48	-627.5***	198.7***	-69.08
	(45.37)	(191.7)	(47.39)	(134.3)
Team FE	YES	YES	YES	YES
Season FE	YES	YES	YES	YES
Observations	117	117	117	117
Number of teams	9	9	9	9

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table 9: Season level IV - First stage - Including clubs always present

	(1)	(2)	(3)	(4)
	Market Value	GHI	Market Value	GHI
Shift Market Value	-0.899**		-2.843***	0.0719
	(0.333)		(0.559)	(0.0464)
Shift GHI		5.015***	70.66***	3.317*
		(0.735)	(13.24)	(1.664)
Constant	15.18***	-4.011***	-44.04***	-2.698*
	(1.502)	(0.712)	(10.39)	(1.411)
Team FE	YES	YES	YES	YES
Season FE	YES	YES	YES	YES
Observations	117	117	117	117
Number of teams	9	9	9	9
R-squared	0.482	0.488	0.661	0.504
F-statistics	7.29	46.5124	14.28	45.19

Robust standard errors in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

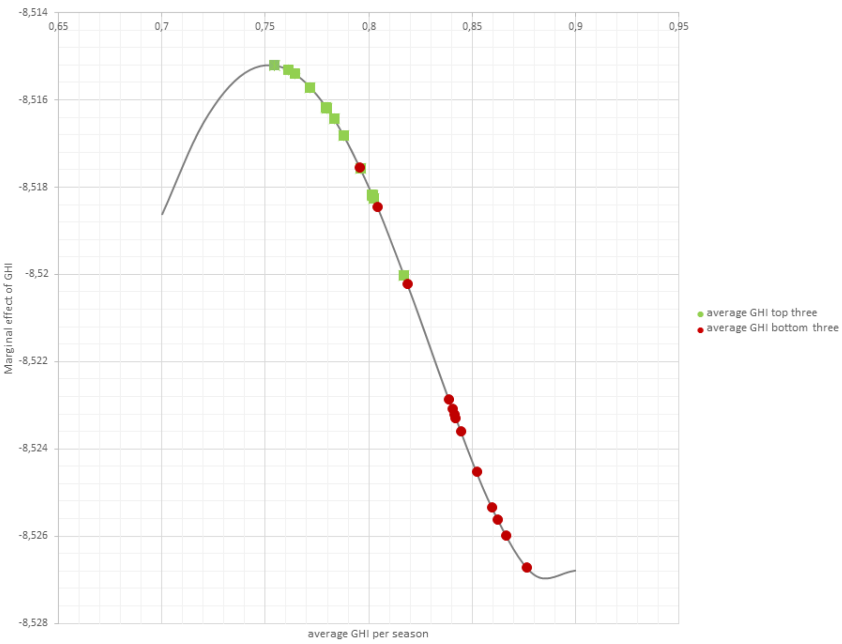

Figure 1: GHI of top three and bottom three teams per season

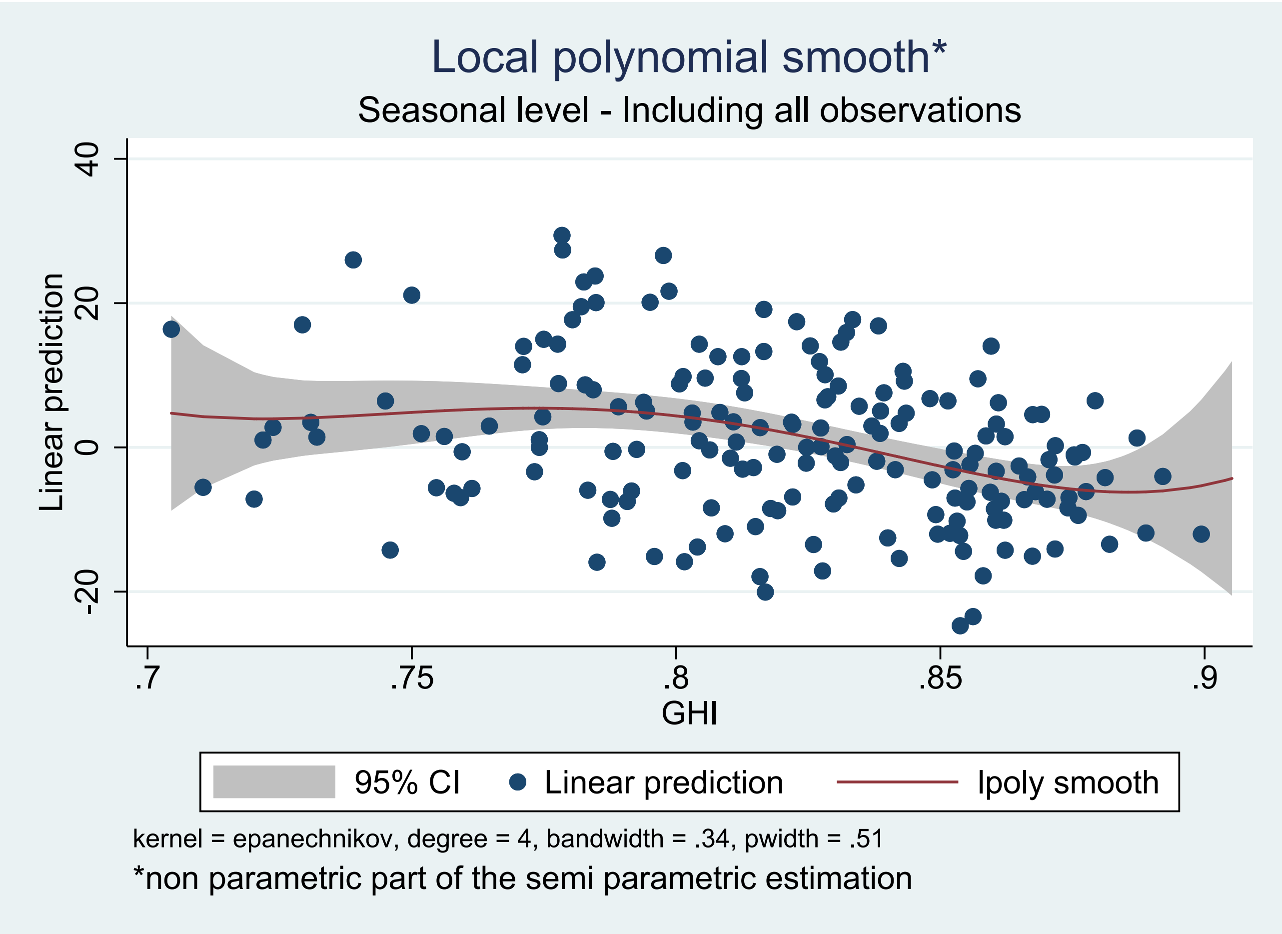

Figure 2: Local polynomial smooth.

Seasonal level - Including all observations

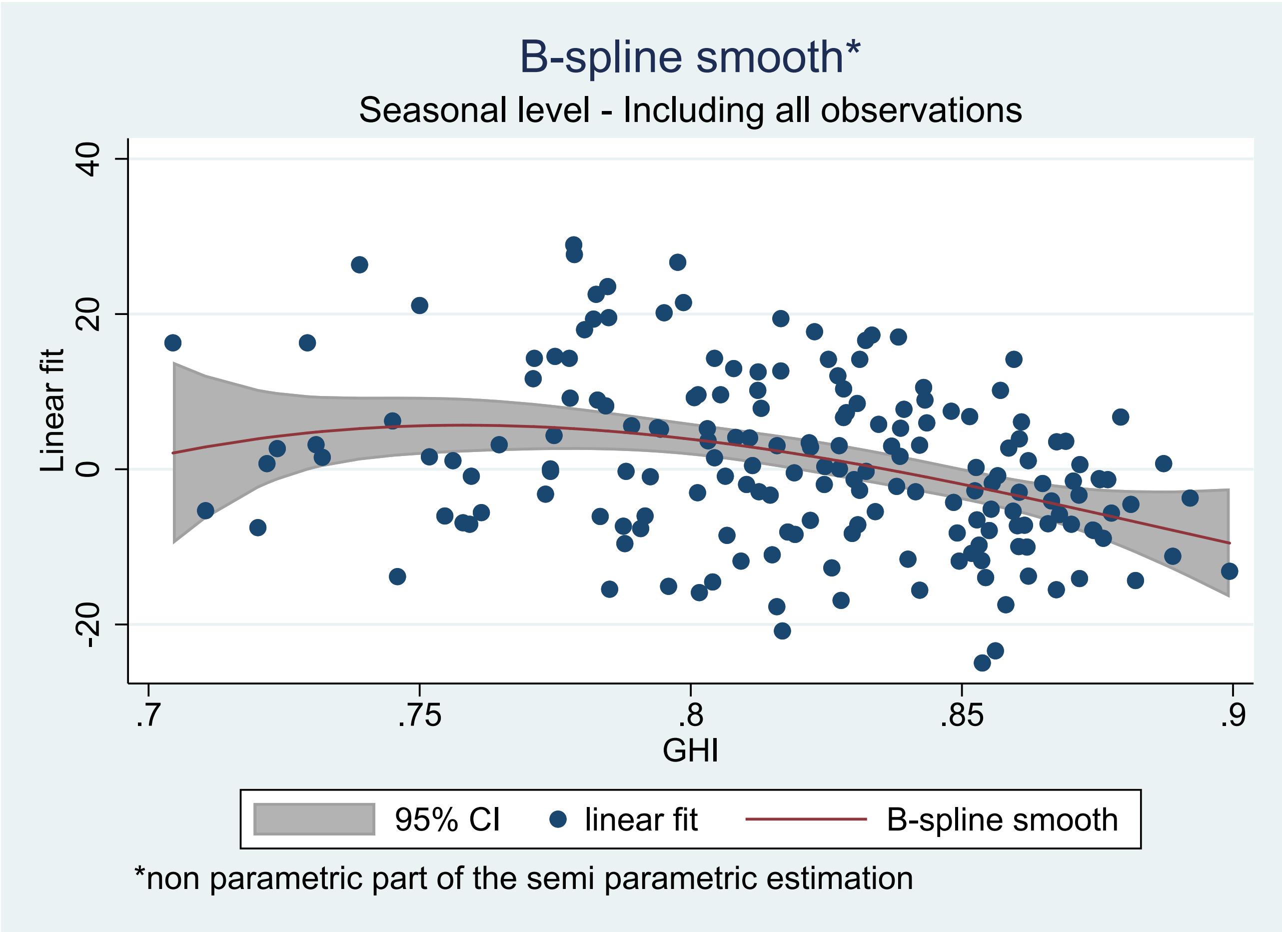

Figure 3: B-spline smooth.

Seasonal level - Including all observations

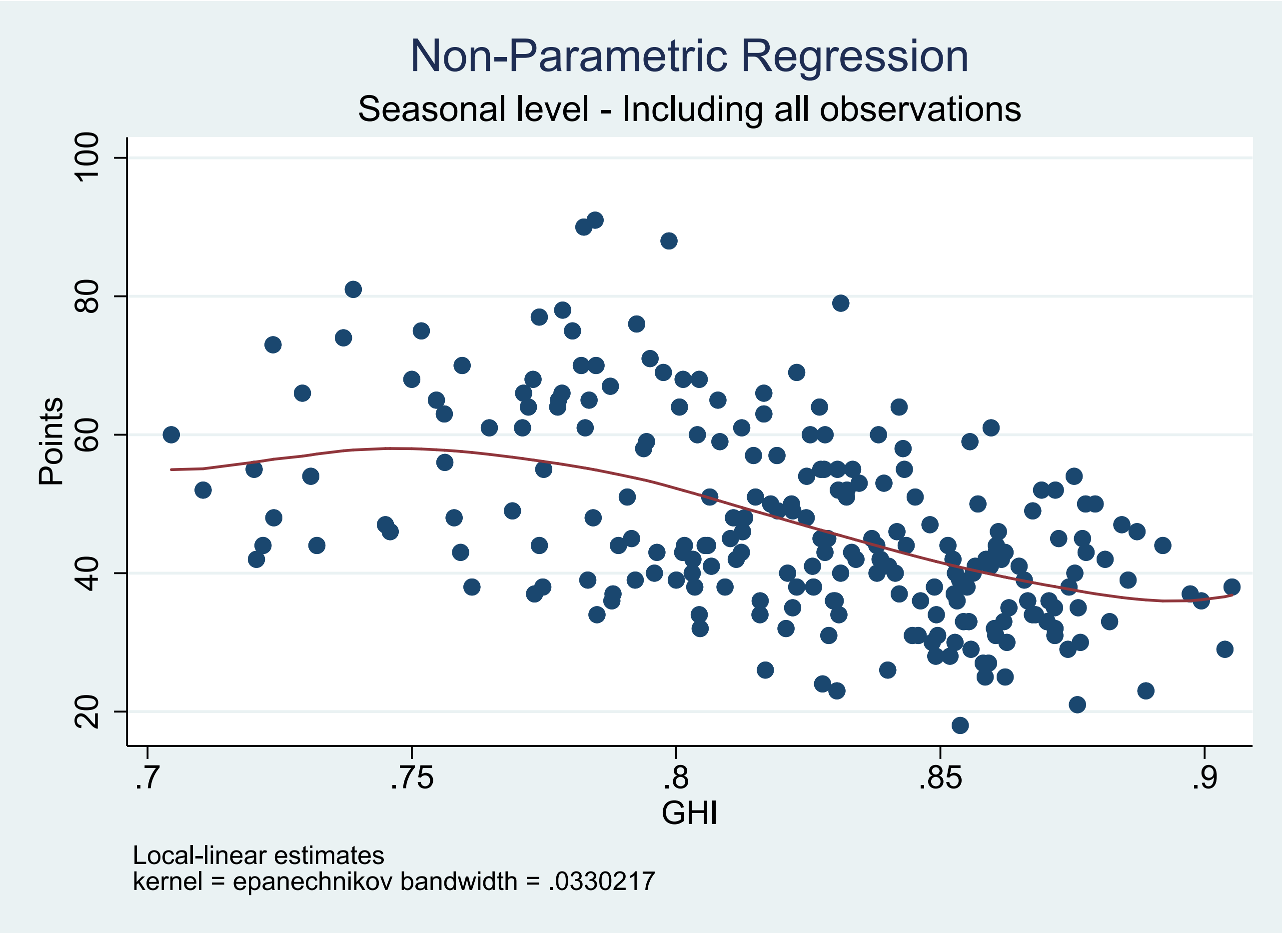

Figure 4: Non-Parametric Regression.

Seasonal level - Including all observations

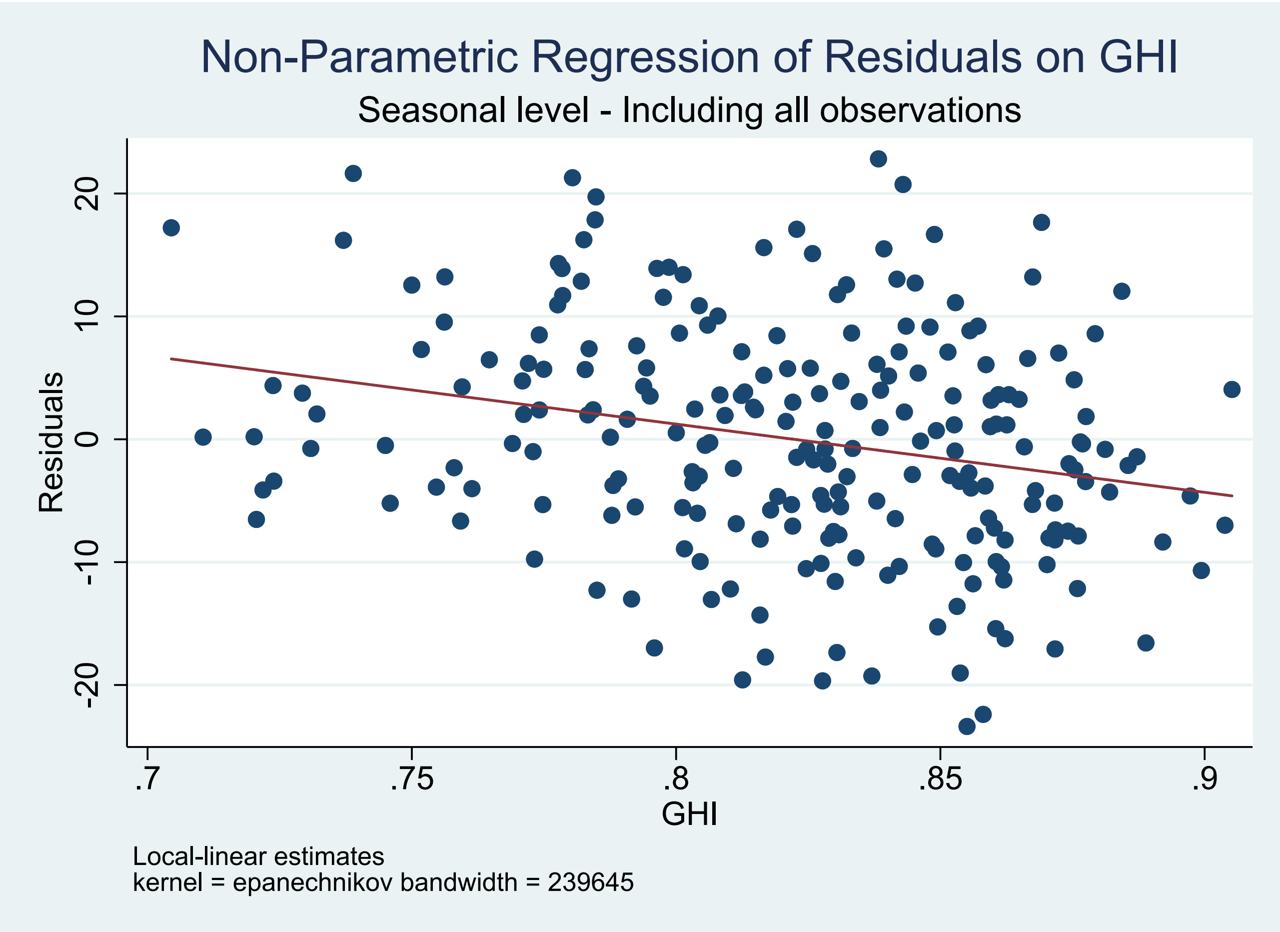

Figure 5: Non-Parametric Regression of Residuals on GHI.

Seasonal level - Including all observations