GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 325

Heterogeneous Contestants and Effort Provision in Tournaments - an Empirical Investigation with Professional Sports Data

* Johannes Berger
** Petra Nieken

* University of Cologne

** University of Bonn

July 2010

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Urs Schweizer. · Department of Economics · University of Bonn · D-53113 Bonn,
Phone: +49(0228)739220 · Fax: +49(0228)739221

Heterogeneous Contestants and Effort Provision in Tournaments - an Empirical Investigation with Professional Sports Data*

Johannes Berger† University of Cologne

Petra Nieken‡ University of Bonn

July 12, 2010

Abstract

We empirically investigate if tournaments between heterogeneous contestants are less intense. To test our hypotheses we use professional sports data from the TOYOTA Handball-Bundesliga, the major handball league in Germany. Using either differences in betting odds or rankings to measure ability differences, our results support standard tournament theory as we find a highly significant negative impact of the matchup's heterogeneity on joint team efforts. However, further analysis shows that this overall decrease in efforts is almost entirely driven by the reaction of the ex-ante favorite team.

Keywords: tournament, heterogeneity, incentives, sports economics
JEL Classification: J24, J33, J41, M52

*We would like to thank Oliver Gürtler, Matthias Krause and the participants of the 13^{th} Colloquium on Personnel Economics in Trier, particularly Bernd Frick, for helpful comments. Petra Nieken gratefully acknowledges financial support of the Deutsche Forschungsgemeinschaft (DFG,) SFB/TR 15.

†Herbert-Lewin-Strasse 2, 50931 Cologne, Germany, e-mail: johannes.berger@uni-koeln.de

‡Adenauerallee 24-42, 53113 Bonn, Germany, email: petra.nieken@uni-bonn.de

1 Introduction

Tournaments, where agents fight for a limited set of given prizes, are omnipresent in day-to-day situations. One can for example observe promotion tournaments or competition for bonus pools within organizations (Baker et al. (1994), Rosen (1986), Rajan and Reichelstein (2006)) or tournaments concerning market shares and litigation contests between them (see for example Taylor (2003), Wärneryd (2000)). We can also model beauty contests, singing contests and sports competitions as tournaments (Amegashie (2009), Szymanski (2003)). For an overview about tournaments and contests see for example Konrad (2009).

As Lazear and Rosen (1981) have shown in their seminal article, rank-order tournaments can -under certain conditions- be the optimal design to induce first best effort levels if only ordinal information is available at reasonable costs. However, theory predicts that incentives are lower in tournaments between heterogeneous contestants than in tournaments between homogeneous ones. In heterogeneous tournaments, the underdog will shy away from competition as his chances of winning are comparably low. The opponent will anticipate this reduction of costly effort and decide to hold back his effort as well. As a result overall performance decreases. This effect is often called the contamination hypothesis. Since in practice contestants are seldom completely homogeneous, this prediction calls the frequent use and effectiveness of tournament schemes in firms and organizations into question. While the logic and effects of heterogeneous tournaments have been studied intensely in the theoretical literature (see among others Kräkel and Sliwka (2004)), only recently a growing body of papers test the theoretical predictions with non-experimental field data from sports contests (for instance Frick et al. (2008), Bach et al. (2009)).[1]

The contribution of this paper to the existing literature is twofold. First,

[1] For experimental evidence see Schotter and Weigelt (1992) or Harbring et al. (2007).

we analyze the impact of heterogeneous tournaments on the provision of team efforts in the TOYOTA Handball-Bundesliga. We are the first to test the contamination hypothesis with data from handball. As we will show, the structure of the game provides a good setting to test this hypothesis. We have collected data of two seasons containing information on goals and fouls as well as ranks and odds from sports betting. Betting odds provide an excellent measure of the team's current ability as they contain all available information such as ranks, injuries or transfers right before each game. We use the number of 2-minute suspensions to approximate the intensity of the game. Our results confirm the contamination hypothesis and show that tournaments between heterogeneous contestants are significantly less intense. The result is robust to different measures of heterogeneity. Further analysis reveals that the overall decrease in effort is almost entirely driven by the reaction of the favorite team to the asymmetry of the match up. Secondly, we provide a direct test if our proxy for team effort is a suitable measure for our analysis. Indeed, we find that the number of 2-minute suspensions positively affects the winning probability of the corresponding team.

Since measures on workers' abilities as well as effort or performance differences would have to be obtained in one dataset, non-experimental field evidence on the contamination hypothesis is quite scarce. Studying professional sports data may help to fill this gap as sports contests often resemble very standardized tournament settings between two parties of which ability and performance proxies may be derived from game statistics. However, the studies which tested the contamination hypothesis with sports data do not provide unambiguous evidence in favor of it. Among the first studies, Ehrenberg and Bognanno (1990) analyze PGA golf tournaments and cannot clearly confirm the contamination hypothesis. They show that the stronger the opponent, the weaker the performance of a player. While this is in line with theory for participants performing below average, it violates theory for participants performing above average as they should be motivated by a higher quality opponent. Brown (2008) also uses data from PGA golf tournaments from 1999-2006 and shows that effort declines if a superstar (Tiger Woods) participates in the tournament. However, her findings are only significant for higher-skilled players but not for lower-skilled ones. Horse race studies like Lynch (2005) support the contamination hypothesis as does Sunde (2009) using tennis data. He also conducts separate analysis for favorites and underdogs and finds that only underdogs are sensitive towards heterogeneity and reduce efforts.

In contrast to our paper all these papers study individual sports. Bach et al. (2009) analyze data from the Olympic Rowing Regatta 2000 for teams and single skulls. They report higher effort levels in homogeneous groups but also find that only the favorites and not the underdogs react to heterogeneity. Bach et al. attribute this finding to the Olympic spirit which might motivate underdogs to do their best irrespective their chances of success. Closest to our paper is the work of Frick et al. (2008) and Nieken and Stegh (2010). Frick et al. use data from the German soccer league. Employing betting odds to measure heterogeneity and red and yellow cards as proxies for effort, their main finding is in line with our results. In this paper, we go one step further and take the dynamic structure of tournaments into account by analyzing the teams' effort responses separately for each half of the game. Our results show that ex-ante ability differences not only determine efforts decisions towards the beginning of the tournament but also towards the end, irrespective the halftime score. Nieken and Stegh analyze the effects of heterogeneity in the German Hockey League. Here the number of minor suspensions also declines if contestants are more heterogeneous. In contrast to our findings, they cannot confirm the contamination hypothesis for each third of the game separately. While we provide evidence that 2-minute suspensions may serve as an effort proxy in the game of handball, all other mentioned work neglect this proof for their data.

The remainder of the paper is structured as follows. The next section describes the dataset and our key variables. In section 3 we present our results and discuss our findings. Section 4 concludes the paper.

2 The Data

In our study we use professional sports data from the first "TOYOTA Handball-Bundesliga", the major handball league in Germany.[2] Our dataset comprises all 612 league games from the seasons 2006/2007 and 2007/2008. For each game and each halftime we collected detailed information on the goals scored and penalties received by either team. We also gathered statistics on the number of spectators, size of venues and the two referees in charge of the game. Even though handball has become the second most popular sport in Germany[3], handball is still rather unknown outside European boarders. For the ease of comprehension the next section will briefly address the most important rules of the game.

2.1 The Very Basics of Handball

In handball[4] two teams, each consisting of one goalkeeper and six field players, compete for two 30 minutes halves. By bouncing, passing and ultimately throwing a small ball into the goal of the opposing team, the team outscoring the opponent wins. In each season all 18 teams play every other team twice, once at home and once away. This amounts to a total of 34 league games for each team in each season. For each game, the winning team earns two championship points while the defeated team receives none. In case of a tie

[2] The data are made publicly available and are downloadable in pdf-format under https:\\www.toyota-handball-bundesliga.de

[3] Among 1046 Germans, 40.7% respondents named handball the second most popular sport after soccer followed by track and field and tennis with roughly 25% and 20% (Statista.de 2009).

[4] Handball is also known as team handball, Olympic handball or European handball. However, American handball is a completely different game.

the two points are split up equally. The championship points determine the final league standing at the end of the season, while the team with the most points wins the national title. In principle all 9 top ranked teams may qualify for a European contest in the upcoming season[5] while up to 3 teams may lose their spot in the first national league. Since almost all final ranks have thus direct implications for the financial future of the ball club, incentives to win are given throughout the entire season.

2.2 Heterogeneous Agents

One of the two key variables needed to test our main hypothesis concerns the heterogeneity of the two agents (teams). Intuitively differences in team abilities should be reflected by the difference in their current league standing. However, this measure may yield poor estimates for some games as standings usually vary substantially at the beginning of each season. Taking the difference in the final ranking may overcome this problem. On the other hand we would then assume constant ability differences over the course of season and ignore potential *ups and downs* caused by injuries or player transfers within the season. A more efficient indicator for the heterogeneity of two teams can be derived from sport betting odds (see Fama (1970), Camerer (1989), Woodland and Woodland (1994), Levitt (2004) or Forrest et al. (2005) for a discussion about market efficiency in betting markets). Betting odds should be able to capture within-seasonal fluctuations of team ability more accurately than rankings. With the odds of a game the implicit winning probabilities of the respective teams are easily calculated (see for instance Deutscher et al. (2009)). To give an example of this calculations consider the game between the TVB Lemgo and HSG Wetzlar which took place on December 12, 2007. Table 1 indicates that the home team TBV

[5] This is the case when German teams have won all three European titles in the previous season as it happened in 2006/2007.

Lemgo was clearly favored by the bookmarkers.[6] The corresponding odds imply that a bettor would receive 1.10 € for every Euro he or she placed on Lemgo. The unlikely case of a tie would yield 13.73 € while a win of the away team would turn every Euro into 7.55 €. From the odds in Table 1 it is straightforward to compute the payout ratio which can then be used to determine the winning probabilities of either team. The payoff ratio is given by the following equation:

$$\frac{1}{\frac{1}{\text{Odd Home Team wins}} + \frac{1}{\text{Odd Tie}} + \frac{1}{\text{Odd Away Team wins}}}$$

Example: 12/29/2007	Betting Odds	Probability
Win of TBV Lemgo	1.10	0.816
Tie	13.73	0.065
Win of HSG Wetzlar	7.55	0.119
Het_Odds	$\|0.816 - 0.119\| = 0.697$	

Table 1: Calculating the Winning Probabilites and Deriving a Heterogeneity Measure from Sports Betting Odds

In the given example the payoff ratio corresponds to 0.8974.[7] Dividing this ratio by the payoffs connected to a win of either home or away team gives the winning probabilities of 0.82 and 0.12 respectively. Taking the absolute difference of these probabilities results is our preferred measure of the match up's heterogeneity: "Het_Odds". This measure can take on any value between 0 (very homogeneous) and 1 (very heterogeneous contestants). As the average in the sample corresponds to 0.49, the contest in the above illustrated example is considered rather heterogeneous by the market.

[6] The odds were retrieved from the sports betting provider betexplorer.de.

[7] The inverse of this ratio corresponds to the average mark-up of the bookmakers which in this case amounts to roughly 10%.

2.3 Penalties as a Measure of Team Effort

The fact that the efforts chosen by workers in firms, teams in handball or by most other agents in non-experimental settings is not directly observable poses an empirical problem when testing the incentive effects of tournaments. Even though in contrast to most firm settings, sports data offer a large amount of statistics, it is not straightforward to decide upon which best reflect individual or team effort. For the game of soccer Frick et al. (2008) argue that team effort is hard to measure with statistics kept on the offensive end of the game. The number of scored goals during a soccer match for instance may not serve as a good proxy for team effort as scoring may also result from a lack of defensive effort by the opposing team. The same argument holds for the game of handball. Similar to Frick et al. (2008), we believe overall team effort is in our case more accurately approximated by the effort put forth in defense which may best be reflected in the team's foul statistics.

Unlike in soccer, a foul in handball is not automatically considered unfair. In general, handball is considered a very physical game, defensive players are allowed to stop the opponent by using body contact when they are in between the attacking player and their own goal. Even though the play is then interrupted and the offensive team regains possession of the ball, such a "fault" is considered a good defensive effort and is not penalized. In fact, if the defensive team can prevent the offense with "faults" from scoring for a long enough time, the referee may eventually call "passive play" urging the offensive team to wrap up the offensive effort. In this case the defense is likely to prevent a goal and get a chance to score themselves on the next possession. Harsher defensive attacks, however, are usually sanctioned by 2-minute suspensions. The player who committed the foul is then temporarily suspended from the game and leaves his team playing a man down for the next 120 seconds.[8] 2-minute suspensions are considered part of the game as

[8] Each player may only receive two 2-minute suspensions. For his third 2-minute sus-

they occur roughly 8 times during an average league game.

In our analysis the sum of 2-minute suspensions will serve as our proxy for joint team effort or, put differently, the overall intensity of the tournament. The idea behind this is as follows: A team who tries particularly hard to prevent the offensive team from scoring will play very physical defense. Sometimes these defensive attacks will be just outside the tolerated norm which should lead to more penalties and in particular more 2-minute suspensions. Since 2-minute suspensions are more frequently ruled than yellow cards in soccer, our effort measure should be less prone to measurement errors such as misjudged referee calls. A novel feature of this paper is that we can provide direct evidence that team penalties indeed reflect overall team effort, as shown in section 3.3.

Table A1 in the appendix provides summary statistics on the committed penalties as well as the main independent variables included in our upcoming analysis.

3 Results

In this section we present our main results. At first we test if the intensity of the game is indeed predicted by the heterogeneity of the particular match up. In section 3.2 we report separate analysis on how ex-ante favorites and ex-ante underdogs react to ability differences in tournaments. Section 3.3 validates our measure of effort by explaining the outcome of the game by the number of 2-minute suspensions ruled against each team.

---

pensions, he automatically receives a red card and is suspended for the rest of the game.

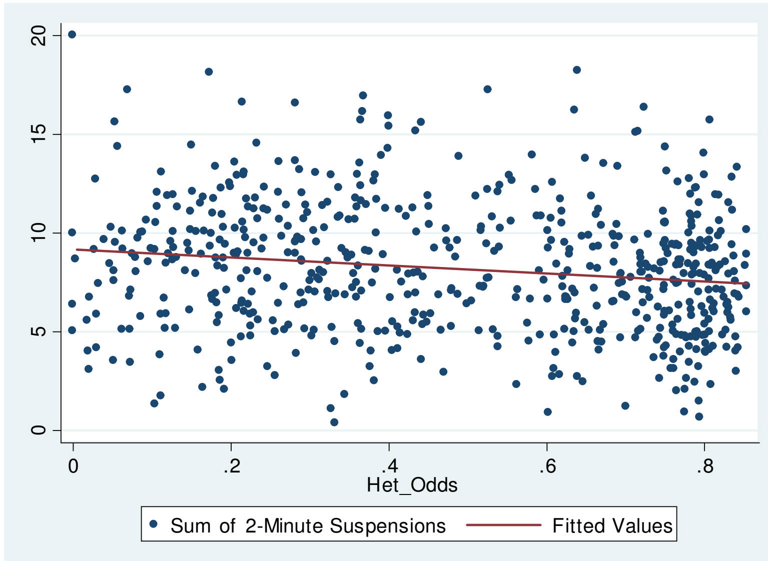

Figure 1: The Relation between Heterogeneity (based on Betting Odds) and the Intensity of the Game

3.1 The Impact of Heterogeneity on the Provision of Efforts

A first descriptive picture on the relation between penalties and ability differences is given in Figure 1. The negative slope of the fitted value line is in support of the contamination hypothesis and reveals that the number of penalties indeed decreases in the heterogeneity of the match up. Of course this conclusion may be shortsighted as it is only based on correlations without any further controls.

To investigate this relation in more detail, we apply regression analysis. As our dependent variable, the sum of 2-minute suspensions, is a count variable we use Poisson regressions throughout our analysis.[9] Our main in-

[9]Figure A1 in the appendix shows that our dependent variable follows a poisson distribution. As shown in Table A1, the variance of our dependent variable is only slighly larger than its mean, indicating that overdispersion is not a problem in our estimations.

dependent variable is the heterogeneity of the two contestants which is approximated by the absolute difference in winning probabilities (Het_Odds) or the absolute difference in final (Het_Final Rank) or current league standings (Het_Current Rank). Besides differences in team abilities, we control for several other factors which are also likely to affect the intensity of a game: As in any other team sport, certain match ups are more important for teams and fans than others. Such games usually take place between two local rivals and are referred to as "derbies". Since these games might in general be fought more intensely, we include a dummy variable (Derby) taking on the value 1 if a game can be classified as a derby and 0 otherwise.[10] Secondly, as pointed out in previous studies, the atmosphere created by fans could affect the players' actions on the court (see for instance Dohmen (2008)). We therefore additionally control for the absolute number of spectators attending the game as well as the percentage of taken seats. Given that some handball venues are much smaller than others, the latter variable gives us a better estimate on how relative attendance affects the intensity of the game. As certain teams might on average be more likely to commit fouls than others, dummy variables for both competing teams are included. To account for the course of the season, dummy variables for season, rounds and their interactions are added. Finally, we also control for referee fixed effects in our estimations.[11]

Table 2 displays our main results. Irrespective of the heterogeneity measure applied, we have highly significant evidence that joint team efforts, as approximated by the sum of 2-minute suspensions per game, are lower in more heterogeneous match ups. Holding all other variables constant, a one standard deviation higher absolute difference in winning probabilities,

However, our results are also robust to other count model specifications as well as simple OLS regressions.

[10] We define a game as a derby if the cities of the two opposing teams are within 150 kilometer distance.

[11] Over the two seasons there were 22 unique pairs of referees.

roughly 26%, is associated with a 6% decrease in the expected sum of 2-minute suspensions.[12] Similar, a one standard deviation larger difference in final standings, roughly 4 ranks, decreases the expected count of penalties by 7%. Furthermore, relative attendance tends to have a positive effect on the intensity of the game. Also penalties are more often ruled in games between two local rivals. However, the two latter results should be interpreted with caution as they are not significant throughout all our specifications.

In Table A2 in the appendix we opt for a nonparametric functional form of our main independent variable to allow for non-linearity of the effect. Here we regress the dependent variable on the 2^{nd} to 5^{th} quintiles of our heterogeneity measures with the lowest quintile of heterogeneity being the reference category in all three specifications. The results clearly support a linear effect of heterogeneity as the number of two minute suspensions steadily decreases in the heterogeneity of match up. Observing a game in the highest quintile of our heterogeneity measure Het_Odds (which on average corresponds to an 80% difference in winning probabilities) as opposed to a game in the lowest quintile of heterogeneity (which on average corresponds to an 13% difference in winning probabilities) decreases the expected count of suspensions by roughly 21%.

In Table A3 we analyze the impact of heterogeneity on effort separately for each half of the game. One could argue that ex-ante ability differences become less important over the course of the game, as the actual difference in goals at halftime provides both teams with a meaningful update of their current ability differences and the respective winning probabilities. However, our results show that even after controlling for the halftime score, ex-ante ability differences play a significant role in predicting the amount of team effort chosen in the last 30 minutes of the game. The insignificant coefficient of "Halftime Score" further indicates that additional information on win-

[12] To compute the percentage change in the expected count of our dependent variable, we use Stata's *listcoef-package* written by J. Scott Long and Jeremy Freese.

Dependent Variable:	Sum of 2-Minute Suspensions		
	(1)	(2)	(3)
Het_Odds	-0.2455***		
	(0.066)		
Het_Final Rank		-0.0192***	
		(0.004)	
Het_Current Rank			-0.0105***
			(0.004)
Derby	0.0777*	0.0634	0.0827**
	(0.040)	(0.041)	(0.041)
Taken seats in %	0.1889*	0.1831*	0.1717
	(0.103)	(0.103)	(0.107)
Spectators/1000	0.0052	0.0041	0.0094
	(0.010)	(0.010)	(0.010)
Constant	1.8065***	1.9099***	1.9406***
	(0.157)	(0.159)	(0.185)
Observations	612	612	594
Pseudo R^2	0.10	0.10	0.10
Log pseudolikelihood	-1415.10	-1412.62	-1377.73

Poisson estimations, robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Further controls: Referees, Teams, Seasons x Rounds (all dummies)

Table 2: The Effect of Heterogeneity on Team Efforts

ning probabilities introduced by performance differences in the first half of the tournament do not seem to affect effort choices in the last part of the tournament.[13]

Overall we believe our results provide rather strong evidence in favor of the contamination hypothesis as predicted by economic theory (Lazear and

[13] Note that "Het_Odds" and "Halftime Score" are highly correlated. However, even if we exclude Het_Odds from the estimation, the difference in goals at the half has only a marginal significant impact on the suspensions ruled in the second half. Also the interaction of the two variables is insignificant.

Rosen (1981)) and confirmed by similar recent empirical studies (e.g. Frick et al. (2008), Bach et al. (2009) or Nieken and Stegh (2010)).

3.2 The Impact of Heterogeneity on Favorites and Underdogs

According to theory, favorites and underdogs[14] should not react differently to the heterogeneity of the match up. In games with heterogeneous contestants, the underdog has only little chances to win and should therefore refrain from providing much effort. The favorite should anticipate this reduction and lower his effort as well. As experimental studies have shown, underdogs often exert higher effort levels than theoretically predicted while the behavior in symmetric settings is roughly in line with theory (see Bull et al. (1987), Schotter and Weigelt (1992)). While Weigelt et al. (1989) find no significant differences when comparing effort levels of favorites and underdogs in unfair tournaments, Harbring and Luenser (2008) report that efforts of weak players are significantly higher than in symmetric settings if the prize spread is high. In a real effort experiment of van Dijk et al. (2001) players with lower ability try to win the tournament against a high ability contestant even though they lose in most cases and could avoid the tournament by playing a piece rate scheme.

Regarding sports data the results are somewhat mixed. While Sunde (2009) show that underdogs react stronger to heterogeneity than favorites, Bach et al. (2009) and Nieken and Stegh (2010) find the opposite. In their studies the favorite lowers his effort in more heterogeneous contests but the efforts of the underdogs remain nearly unchanged. One may argue that in sports the general norm suggests not to give up irrespective the size of the deficit. In team sports this norm might be even more prominent as players do not want to let their teammates and coaches down. From an individual

[14] We define favorites and underdogs according to the betting odds for each game.

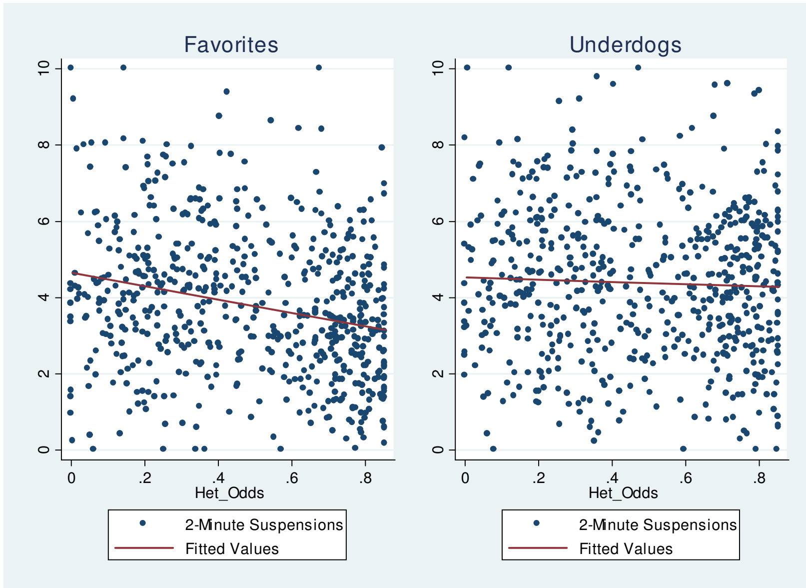

Figure 2: The Relation between Heterogeneity (according to Betting Odds) on Team Effort of Favorites and Underdogs

player's perspective, giving up could also result in being put to the bench in the next game. In contrast, the favorite team may lower its effort without having to fear such social sanctions as long as it wins the game which is the case in 75% of the sampled games. We therefore expect favorites to be more willing to withhold efforts in heterogeneous contests than underdogs.

In Figure 2 we show a scatter plot of committed 2-minute suspensions and the heterogeneity of the match up (Het_Odds) separately for favorites and underdogs. The picture seems to support the results found in Bach et al. (2009) and Nieken and Stegh (2010) as the favorites' efforts are substantially lower in heterogeneous contests while the right panel of Figure 2 shows no systematic pattern for the underdogs. The overall decrease in the games' intensities previously shown in Figure 1 thus seems to be driven by the adjustments of the ex-ante favorites.

To confirm this impression we run separate regressions for favorites and underdogs explaining the teams' committed penalties by the heterogeneity of the match up. Except for the dependent variable, the specifications in columns (1) and (3) are identical to our previous specification in Table 2. In specification (2) and (4) we additionally test the linearity of the effect by regressing our dependent variable on the quintiles of our heterogeneity measure. The results show that indeed only the ex-ante favorite reacts to heterogeneity by reducing effort. The coefficient in column (1) suggests that a one standard deviation increase in our measure "Het_Odds" reduces the expected count of 2-minute suspension of the favorite team by roughly 10%. Column (2) indicates that the favorite's reaction to the heterogeneity is linear as indicated by the growing economic and statistical significance of higher quintile coefficients. On average, the favorite's expected penalties are about 32% lower when the difference in the ex-ante winning probabilities falls into the 5^{th} quintile as opposed to the 1^{st} quintile. Interestingly the coefficient for "Derby" is highly significant in both estimations, suggesting that the favorite is willing to sacrifice additional effort when playing against one of his rivals. Columns (3) and (4) reveal that the underdog's effort is hardly affected by ex-ante ability differences as all coefficients are economically and statistically insignificant. Table A4 shows that this result is also reflected in the raw means of the dataset.

This finding is in line with Bach et al. (2009) and Nieken and Stegh (2010) but stands in sharp contrast to standard tournament theory. As mentioned above this result may be attributed to some social costs faced by inferior contestants for giving up. A similar argument is brought forward in a recent study by Fershtman and Gneezy (2010). In their field experiment the majority of students participating in running tournaments are unwilling to quit or drop out of the contest even when their prospects to win become negligible.

Some suggestive evidence for the existence of social sanctions imposed by the fans comes from Table A5. Here we run separate regressions explaining

Dependent Variable:	2-Minute Suspensions			
	Favorite		Underdog	
	(1)	(2)	(3)	(4)
Het_Odds	-0.4063***		0.0977	
	(0.099)		(0.094)	
2^{nd} Quintile		-0.0038		-0.0717
		(0.057)		(0.049)
3^{rd} Quintile		-0.1056*		-0.0028
		(0.056)		(0.054)
4^{th} Quintile		-0.1937***		0.0211
		(0.070)		(0.058)
5^{th} Quintile		-0.3838***		-0.0200
		(0.081)		(0.072)
Derby	0.1580***	0.1549***	0.0517	0.0487
	(0.058)	(0.058)	(0.050)	(0.050)
Seats taken in%	0.1253	0.1189	0.1175	0.1210
	(0.114)	(0.112)	(0.103)	(0.104)
Spectators/1000	-0.0171*	-0.0147	0.0014	0.0046
	(0.010)	(0.010)	(0.008)	(0.008)
Constant	0.8979***	0.8382***	1.2510***	1.2629***
	(0.246)	(0.254)	(0.196)	(0.193)
Observations	612	612	612	612
Pseudo R^2	0.09	0.09	0.08	0.09
Log pseudolikelihood	-1132.67	-1129.70	-1165.08	-1164.45

Poisson estimations, robust standard errors in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1

Further controls: Referees, Teams, Seasons x Rounds (all dummies)

Table 3: The Effect of Heterogeneity on Team Effort of Favorites and Underdogs

the intensity put forth by the favorite during home and away games. The coefficient of "Het_Odds" in column (3) shows that the effort reduction in heterogeneous matches is particularly large when the favorite does not play in front of the home crowd. Especially games in the highest quintile of heterogeneity are played significantly less intensely when the game is played away as suggested by our estimation in column (4). However, the difference in effort reduction at away games is not quite significant when introducing an interaction term into a pooled estimation of both, home and away games.

3.3 Testing our Measure of Effort

How do we know that the number of 2-minute suspensions really serves as a good measure of team effort? On average more effort should increase a team's probability to win. If the number of suspensions truly reflects the amount of defensive and thus team effort, more suspensions should be positively associated with the team's probability to win. On the other hand, one could also expect to see a decrease in the winning probability as the team has to play a man down whenever a suspension is ruled.[15] To validate our measure of effort Table 4 explains the outcome of the game by the share of penalties ruled against the ex-ante favorite team.[16] In specifications (1-4) our dependent variable is the difference in goals, i.e., the goals scored by the favorite team minus the goals scored by the underdog, while in columns (5-8) explain the probability for the favorite team to win. If our line of thought is correct, an increase in the share of 2-minute suspensions should lead to a more favorable outcome for the corresponding team. This reasoning is partially confirmed in column (1) in which we explain the difference in scored goals by a simple OLS regression. Controlling for the ex-ante winning probabilities and team fixed

[15]However, results from soccer for instance indicate that even the permanent expulsion of a player does not necessarily lead to a disadvantage for the affected team (e.g. Caliendo and Radic (2006)).

[16]Note that the denominator of this measure already accounts for the overall intensity of the game as well as the number of fouls committed by the underdog.

effects, the share of 2-minutes suspensions ruled against the favorite team has a positive and marginally significant impact on the difference in goals. In column (2) we again test for the linearity of this effect and see that the best outcome is achieved when the share of penalties rises to the 4^{th} quintile. In specifications (3) and (4) we repeat the previous estimations but restrict our sample to the 50% most homogeneous games. In these games a team's marginal effort should have the largest impact on the outcome of the game. In the remaining games, ex-ante ability differences may be so large that the teams can hardly affect the outcome by increasing or decreasing efforts. Indeed we find a much stronger and highly significant effect of the share of penalties on the difference in scored goals among homogeneous contestants.

However, a team's effort should primarily be directed toward winning the game and not toward outscoring the opponent by many goals. A more appropriate way to validate our effort measure is therefore to test its direct impact on the team's winning probability. Again a simple descriptive statistic seems to suffice to support our argument. In the games which were won by the favorite the average share of suspensions ruled against the favorite amounts to 46.35% while in the games that were lost this number corresponds to 44.48%. In specification (5-8) of Table 4 we further test this difference by regressing a dummy variable taking on the value 1 if the favorite team wins and 0 otherwise on the share of penalties and the control variables used in the previous specification. In specification (5-6) we again include all games in the analysis while (7-8) includes only the most homogeneous games. The displayed coefficients are the marginal effects from our probit regression. Again the share of 2-minute suspensions ruled against the favorite has a significant positive effect on the winning probability. The coefficient in column (6) implies that a 10% rise in the share of 2-minute suspensions increases the likelihood of a win on average by 2.6%. In homogeneous games this effect is about 3 times as large.[17] The results in column (6) and (8)

[17] Among the 50% most homogeneous games, the favorite committed 51% of the penalties

again imply that the effect of the share of 2-minute suspensions is more or less linear. The coefficient for the 5^{th}quintile, however, indicates that committing too many penalties may eventually reverse this positive effect. Summing up, Table 4 provides direct evidence that number of 2-minute suspensions indeed reflect the effort put forth by a handball team. This result is reassuring for our main findings reported in the previous two subsections.

in the games he won and only 44% in the games which were lost.

Dependent Variable:	Goals Favorite -Goals Underdog				Favorite Wins			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Games		Low Heterogeneity		All Games		Low Heterogeneity	
2- Minutes Share Favorite	0.0252*		0.0623***		0.0026**		0.0066***	
	(0.013)		(0.018)		(0.001)		(0.002)	
2^{nd} Quintile		0.9664		2.0440*		0.0364		0.1266
		(0.640)		(1.048)		(0.049)		(0.092)
3^{rd} Quintile		0.2933		1.7235*		0.0799*		0.2066**
		(0.580)		(0.920)		(0.046)		(0.084)
4^{th} Quintile		1.6038**		2.5448**		0.1101**		0.3102***
		(0.789)		(1.133)		(0.050)		(0.072)
5^{th} Quintile		1.2072*		2.7772***		0.0726		0.2063**
		(0.679)		(0.906)		(0.049)		(0.086)
Taken seats in %	-0.0373	-0.0448	1.0113	0.7554	0.0997	0.0914	0.1298	0.0576
	(1.120)	(1.131)	(1.832)	(1.840)	(0.115)	(0.116)	(0.205)	(0.208)
Spectators/1000	0.0733	0.0859	-0.1777	-0.1549	-0.0147*	-0.0143*	-0.0310*	-0.0287*
	(0.094)	(0.093)	(0.139)	(0.138)	(0.009)	(0.009)	(0.016)	(0.016)
Het_Odds	8.4659***	8.3464***	6.5613***	6.6384***	0.5295***	0.5171***	0.6046**	0.6732**
	(1.158)	(1.169)	(2.270)	(2.303)	(0.099)	(0.099)	(0.260)	(0.263)
Constant	-5.4164**	-5.1116*	-7.7445**	-6.6801*				
	(2.714)	(2.674)	(3.624)	(3.683)				
Observations	611	611	305	305	611	611	305	305
R^2 or Pseudo R^2	0.34	0.34	0.22	0.22	0.20	0.20	0.15	0.15
Log pseudolikelihood					-276.97	-276.80	-174.91	-173.91

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, robust standard errors in parentheses

(1-4) OLS estimations, (5-8) Probit estimation (marginal effects reported), Further controls: Teams (dummies)

Table 4: Team Effort as a Determinant for the Outcome of the Game

4 Conclusion

Organizations often implement tournament schemes to induce incentives and decide about promotions of their employees. Indeed, tournaments can lead to first best effort levels but effort is predicted to decline if contestants are heterogeneous. Since in reality contests are seldom completely homogeneous, the effectiveness of tournaments in practice is called into question. As our analysis has shown, there is strong evidence in favor of the contamination hypothesis, i.e., heterogeneity between teams leads to a less intensive tournament. Especially the ex-ante favorite is likely to withhold effort while the underdog does not cease to exert effort "against all odds". In the game of handball or in team sports in general the latter result may be explained by some social or psychological costs the inferior contestant faces when not trying hard enough against an ex-ante dominant rival. However, in organizations such social costs may be absent or considerably lower as effort provision is not as publicly observable as it is in sports. In organizations underdogs might therefore also decide to spare costly effort when the prospects to win are considerably low. To prevent this overall decrease in performance, firms should try to set up tournaments between contestants of similar ability. While in sports relegation systems or payroll caps help to ensure a competitive balance, firms can, for instance, match constants with equal job profiles, educational background or tenure. If this is not possible, firms may have to handicap the more able contestant (see for instance Lazear and Rosen (1981) or Knoeber and Thurman (1994)), add absolute performance standards or refrain from using tournaments schemes at all.

References

Amegashie, J. A. (2009). American idol: Should it be a singing contest or a popularity contest? *Journal of Cultural Economics 33*, 265–277.

Bach, N., O. Guertler, and J. Prinz (2009). Incentive effects in tournaments with heterogeneous competitors - an analysis of the olympic rowing regatta in sydney 2000. *Management Revue. The International Review of Management Studies 20*(3), 239–253.

Baker, G., R. Gibbons, and K. J. Murphy (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics 109*, 1125–56.

Brown, J. (2008). Qitters never win: The (adverse) incentive effects of competing with super stars. Working Paper.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy 95*, 1–33.

Caliendo, M. and D. Radic (2006). Ten do it better, do they? an empirical analysis of an old football myth. *IZA Discussion Paper No. 2158*.

Camerer, C. F. (1989). Does the basketball market believe in the 'hot hand,'? *The American Economic Review 79*(5), 1257–1261.

Deutscher, C., B. Frick, O. Gürtler, and J. Prinz (2009). Sabotage in heterogeneous tournaments: A field study. *Working Paper*.

Dohmen, T. J. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic Inquiry 46*, 411–424.

Ehrenberg, R. G. and M. L. Bognanno (1990). Do tournaments have incentive effects? *Journal of Political Economy 98*, 1307–1324.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*, 383–417.

Fershtman, C. and U. Gneezy (2010). The trade-off between performance and quitting in high-power tournaments. *Journal of the European Economic Association forthcoming*.

Forrest, D. J., J. Goddard, and R. Simmons (2005). Odds-setters as forecasters: The case of english football. *International Journal of Forecasting 21*, 551–564.

Frick, B., O. Gürtler, and J. Prinz (2008). Anreize in turnieren mit heterogenen teilnehmern: Eine empirische untersuchung mit daten aus der fussball-bundesliga. *Zeitschrift für betriebswirtschaftliche Forschung 60*, 385–405.

Harbring, C., B. Irlenbusch, M. Kräkel, and R. Selten (2007). Sabotage in corporate contests - an experimental analysis. *International Journal of the Economics of Business 14*, 367–392.

Harbring, C. and G. Luenser (2008). On the competition of asymmetric agents. *German Economic Review 9*(3), 373 – 395.

Knoeber, C. and W. Thurman (1994). Testing the theory of tournaments: An empirical analysis of broiler production. *Journal of Labor Economics 12*, 155–179.

Konrad, K. (2009). *Strategy and Dynamics in Contests*. Oxford Business Press.

Kräkel, M. and D. Sliwka (2004). Risk taking in asymmetric tournaments. *German Economic Review 5*, 103–116.

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

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets?. *Economic Journal 114*(495), 223–246.

Lynch, J. (2005). The effort effects of prizes in the second half of tournaments. *Journal of Economic Behavior & Organization 57*, 115–129.

Nieken, P. and M. Stegh (2010). Incentive effects in asymmetric tournaments - empirical evidence from the german hockey league. SFB TR 15 Discussion Paper No. 305.

Rajan, M. V. and S. Reichelstein (2006). Subjective performance indicators and discretionary bonus pools. *Journal of Accounting Research 44*(3), 585–618.

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

Schotter, A. and K. Weigelt (1992). Asymmetric tournaments, equal opportunity laws, and affirmative action: some experimental results. *Quarterly Journal of Economics 107*, 511–539.

Sunde, U. (2009). Heterogeneity and performance in tournaments: a test for incentive effects using professional tennis data. *Applied Economics 41*, 3199 – 3208.

Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature 41*, 1137–1187.

Taylor, J. (2003). Risk-taking behavior in mutual fund tournaments. *Journal of Economic Behavior & Organization 50*, 373–383.

van Dijk, F., J. Sonnemans, and F. van Winden (2001). Incentive systems in a real effort experiment. *European Economic Review 45*, 187–214.

Weigelt, K., J. Dukerich, and A. Schotter (1989). Reactions to discrimination in an incentive pay compensation scheme: A game-theoretic approach. *Organizational Behavior and Human Decision Processes 44*, 26–44.

Woodland, L. M. and B. M. Woodland (1994). Market efficiency and the favorite-longshot bias: The baseball betting market. *The Journal of Finance 49*(1), 269–279.

Wärneryd, K. (2000). In defense of lawyers: Moral hazard as an aid to cooperation. *Games and Economic Behavior 33*(1), 145–158.

5 Appendix

Label	*Description*	*Mean*	*SD*	*Min*	*Max*
Round	Round of Season	17.5	9.82	1	34
Referees	Id of Referees	9.25	5.47	1	22
Goals	Sum of Scored Goals	59.18	7.49	38	87
Goals Favorite − Goals Underdog	Goals of the Favorite − Goals of the Underdog	4.22	5.65	-15	27
Favorite Wins	Dummy=1 if Favorite wins and 0 otherwise	0.75	0.44	0	1
Sum of 2-Minute Suspensions	2-Minute Suspensions per Game	8.18	3.16	0	19
2-Minutes Favorite	2-Minute Suspensions of the Favorite	3.79	1.89	0	10
2-Minutes Underdog	2-Minute Suspensions of the Underdog	4.49	1.94	0	10
2- Minutes Share Favorite	2-Minutes Favorite / 2-Minute Suspensions	45.88	15.91	0	100
Het_Odds	Abs. Diff in Winning Probabilities	0.49	0.26	0.00	0.85
Het_Final Rank	Abs. Diff in Standings at the End of the Season	6.33	4.11	1	17
Het_Current Rank	Abs. Diff in Current Standings	6.09	3.96	1	17
Halftime Score (Diff)	Abs. Diff in Goals at Halftime	3.76	2.83	0	13
Taken seats in %	Attendance Relative to Venue Size	0.78	0.20	0.11	1
Spectators/1000	Spectators per Game divided by 1000	4.76	3.11	1.00	19.40
Derby	Dummy=1 if Cities of Teams are within 150km Distance	0.13	0.33	0	1

Table A1: Descriptive Statistics of Key Variables

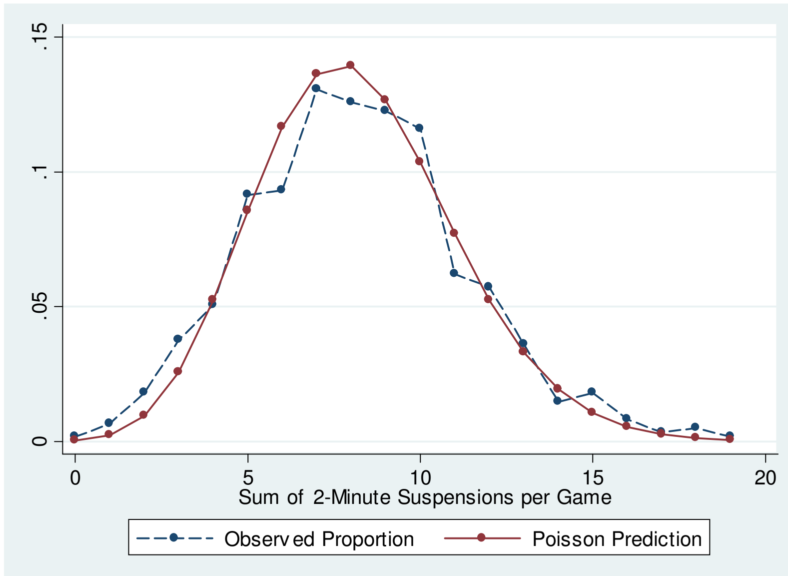

Figure A1: The Distribution of 2-Minute Suspensions in the Sample

Dependent Variable:	Sum of 2-Minute Suspensions		
	(1)	(2)	(3)
	Het_Odds	Het_Final Rank	Het_Current Rank
2^{nd} Quintile	-0.0477	-0.0681*	0.0095
	(0.040)	(0.039)	(0.041)
3^{rd} Quintile	-0.0786*	-0.1242***	-0.0838**
	(0.043)	(0.039)	(0.042)
4^{th} Quintile	-0.1237***	-0.1826***	-0.0972**
	(0.047)	(0.044)	(0.045)
5^{th} Quintile	-0.2410***	-0.2196***	-0.1146**
	(0.054)	(0.052)	(0.049)
Derby	0.0706*	0.0631	0.0855**
	(0.040)	(0.040)	(0.041)
Taken seats in %	0.2230**	0.1807*	0.1642
	(0.103)	(0.102)	(0.106)
Spectators/1000	0.0024	0.0033	0.0086
	(0.010)	(0.010)	(0.010)
Constant	1.7722***	1.8909***	1.9379***
	(0.159)	(0.156)	(0.182)
Observations	612	612	594
Pseudo R^2	0.10	0.10	0.10
Log pseudolikelihood	-1412.71	-1411.08	-1376.10

Poisson estimations, robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Further controls: Referees, Teams, Seasons x Rounds (all dummies)

Table A2: Heterogeneity and Team Efforts

Dependent Variable:	Sum of 2-Minute Suspensions			
	1^{st} Half		2^{nd} Half	
	(1)	(2)	(3)	(4)
Het_Odds	-0.0207**		-0.0186***	
	(0.008)		(0.007)	
Het_Final Rank		-0.0212***		-0.0162**
		(0.008)		(0.007)
Derby	0.0531	0.0429	0.0717	0.0660
	(0.074)	(0.074)	(0.061)	(0.061)
Seats taken in %	0.0853	0.0596	0.3173*	0.2864*
	(0.215)	(0.215)	(0.174)	(0.174)
Spectators/1000	-0.0023	-0.0016	0.0044	0.0065
	(0.019)	(0.019)	(0.014)	(0.014)
Halftime Score (Diff)			-0.0075	-0.0079
			(0.008)	(0.008)
Constant	0.9312***	1.0838***	1.1994***	1.3056***
	(0.334)	(0.344)	(0.278)	(0.286)
Observations	611	611	611	611
Pseudo R^2	0.09	0.09	0.08	0.08
Log pseudolikelihood	-1116.43	-1115.88	-1261.622	-1262.15

Poisson estimations, robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Further controls: Referees, Teams, Seasons x Rounds (all dummies)

Table A3: Heterogeneity and Team Efforts in Each Half

	2- Minute Suspensions (Game Averages)		
Het_Odds	Sum	Favorite	Underdog
1^{st} Quintile	8.73	4.18	4.55
2^{nd} Quintile	8.93	4.42	4.52
3^{rd} Quintile	8.11	3.84	4.28
4^{th} Quintile	7.94	3.64	4.30
5^{th} Quintile	7.16	2.88	4.29
All Games	8.18	3.79	4.39

Table A4: The Relation between Heterogeneity and Effort

Dependent Variable:	2-Minute Suspensions Favorite			
	Favorite is Home Team		Favorite is Away Team	
	(1)	(2)	(3)	(4)
Het_Odds	-0.2506		-0.4920**	
	(0.215)		(0.212)	
2^{nd} Quintile		0.0031		0.0102
		(0.075)		(0.087)
3^{rd} Quintile		-0.1575*		-0.1314
		(0.091)		(0.103)
4^{th} Quintile		-0.1408		-0.2031
		(0.117)		(0.129)
5^{th} Quintile		-0.2687*		-0.4470***
		(0.141)		(0.173)
Derby	0.2676***	0.2649***	-0.0749	-0.0984
	(0.079)	(0.078)	(0.100)	(0.102)
Seats taken in %	-0.0931	-0.0719	0.0689	-0.0029
	(0.158)	(0.157)	(0.263)	(0.260)
Spectators/1000	0.0018	-0.0010	0.0262	0.0439
	(0.015)	(0.015)	(0.028)	(0.029)
Constant	1.1922***	1.1620***	1.3156***	1.2675***
	(0.152)	(0.144)	(0.292)	(0.286)
Observations	404	404	208	208
Pseudo R^2	0.07	0.07	0.06	0.06
Log pseudolikelihood	-765.08	-763.33	-395.41	-394.29

Poisson estimations, robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Further controls: Teams (all dummies)

Table A5: The Effect of Heterogeneity on the Favorites' Effort when Playing Away