



**HAL**  
open science

## Three essays in applied economics

Jean-Baptiste Vilain

► **To cite this version:**

Jean-Baptiste Vilain. Three essays in applied economics. Economics and Finance. Institut d'études politiques de paris - Sciences Po, 2018. English. NNT : 2018IEPP0009 . tel-03419493

**HAL Id: tel-03419493**

**<https://theses.hal.science/tel-03419493>**

Submitted on 14 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Institut d'études politiques de Paris  
ECOLE DOCTORALE DE SCIENCES PO  
Programme doctoral en économie  
Département d'économie

Doctorat en sciences économiques

## **Three essays in applied economics**

**Jean-Baptiste VILAIN**

Thèse dirigée par Alfred GALICHON et  
supervisée par Philippe FÉVRIER  
défendue le 15 juin 2018

### **Jury**

- Luc ARRONDEL, Directeur de recherche CNRS, Paris School of Economics  
*Rapporteur*
- Ghazala AZMAT, Professeur des Universités, IEP de Paris
- Philippe FÉVRIER, Professeur associé, Ecole Polytechnique & CEO de Veltys  
*Responsable scientifique au sein de l'entreprise Veltys (thèse CIFRE)*
- Alfred GALICHON, Professor of Economics and of Mathematics, New York University  
*Directeur de thèse*
- Michael VISSER, Directeur de recherche CNRS, CREST  
*Rapporteur*



# Remerciements

Je souhaite en premier lieu remercier chaleureusement mon directeur de thèse académique, Alfred Galichon. Alfred est un chercheur que j'admire beaucoup et son cours *Matching markets*, que j'ai suivi en Master, a été déterminant dans mon choix de me lancer dans une thèse en économie. J'ai particulièrement apprécié la confiance qu'il a su m'accorder pour le choix de mes thématiques de recherche, d'autant que celles-ci ont considérablement évolué par rapport à mon projet initial. Je lui suis également reconnaissant pour son soutien et ses précieux conseils tout au long de la thèse. Ce fut un honneur et une chance de travailler sous sa direction.

Je tiens aussi à exprimer toute ma gratitude envers mon directeur côté entreprise, Philippe Février. Nos échanges, très stimulants, m'ont toujours permis de très bien avancer, à la fois sur les points techniques, la structuration des raisonnements et l'organisation des idées. De plus, j'ai particulièrement apprécié le fait qu'il soit exigeant tout en ayant confiance en moi et en étant bienveillant et positif. Après mes points avec lui, je me sentais très motivé avec l'envie d'aller de l'avant. J'ai eu beaucoup de chance de croiser son chemin et je suis très heureux et fier de continuer à travailler avec lui chez Veltys, maintenant que ma thèse se termine.

Je remercie aussi Luc Arrondel et Michael Visser d'avoir accepté d'être rapporteurs, et Ghazala Azmat de participer au jury. Je garde d'excellents souvenirs de mes échanges avec Ghazala et Luc qui m'ont permis de beaucoup améliorer les chapitres 2 et 3.

Mes coauteurs, avec des styles et des personnalités très différents, ont aussi joué un rôle très important dans le déroulement de ma thèse. Ce fut un grand plaisir de travailler avec chacun d'entre eux. Je remercie Laurent Rossignol pour sa connaissance du terrain et ses blagues potaches, Antoine Chapsal pour ses idées foisonnantes et son optimisme, et Rodrigo Lopez-Kolkovsky pour m'avoir donné envie de découvrir les méthodes de machine learning. Ma thèse n'aurait pas eu la même saveur si je n'avais pas eu la chance de travailler et de passer de très bons moments avec ce trio.

Je dois aussi beaucoup aux interactions que j'ai eu la chance d'avoir avec certains chercheurs, notamment Wladimir Andreff, Richard Duhautois, Gabrielle Fack et Lars Persson. Aller présenter mes papiers lors de séminaires ou de conférences et échanger de façon plus informelle avec des enseignants chercheurs est certainement l'une des facettes de la thèse que j'ai préférée.

Je remercie aussi les doctorants de Sciences Po, en particulier Pierre Cotterlaz, Pierre Deschamps, Marcos Diaz, Etienne Fize, Arthur Guillouzouic-Le Corff (merci pour les petits cours de micro théorique!), Charles Louis-Sidois, Andreea Minea, Jan Sonntag et Lucas Vernet, avec qui j'ai passé de très bons moments. Du côté de Veltys, je suis très reconnaissant envers Violette Nahmias pour m'avoir convaincu que la meilleure des analyses possibles ne valait rien si elle n'était pas accompagnée de son "histoire", et envers Romain Aeberhardt pour ses retours sur mes papiers qui m'ont souvent aidé à y voir plus clair. Je remercie aussi tous les salariés et stagiaires de Veltys qui rendent l'ambiance de travail dans cette entreprise très agréable avec un clin d'oeil tout particulier à mes anciennes co-bureaux MAPP Mariya Litinskaya et Anastasia Shchepetova et à la Team Upply. Merci aussi aux équipes administratives qui ont très bien géré la complexité organisationnelle d'une thèse CIFRE : Brigitte Contri chez Veltys, Claudine Lamaze et Pilar Calvo à Sciences Po, Fabienne Roussie à l'ANRT.

Mes remerciements vont aussi à mes proches et à mes amis : Saïd Kamel pour m'avoir encouragé à faire de la recherche; Laure Mercier pour ses corrections de langue sur le chapitre 3; Nicolas Cazy, Nicolas Kisgen, Thibaut Griffart et Léo Simonin pour m'avoir changé les idées et remotivé tout au long de la thèse; Eloy Lafaye, Clément Leroy et les PIAF pour tous les beaux moments partagés ensemble.

Je souhaite évidemment remercier du fond du coeur mes parents pour leur soutien indéfectible dans tous mes projets personnels et professionnels. J'ai vraiment énormément de chance de vous avoir et je vous dois *"jusqu'à ce qu'on ne puisse plus compter"*.

Enfin, mes pensées les plus tendres vont à Aude, qui doit se réjouir que j'arrive au bout de cette thèse. Merci d'avoir écouté et corrigé mes présentations avant chaque séminaire/conférence, de m'avoir constamment encouragé et de m'avoir supporté pendant le dernier mois de la thèse. Surtout merci d'être toi, je suis si heureux de devenir ton mari cet été...

# Notice

The three chapters of this dissertation are self-contained research articles. Therefore, the term “paper” or “article” are used and parts of the content may be repeated. The three chapters are coauthored, which explains the use of the “we” pronoun.



# Contents

Summary	11
<b>1 The impact of teachers on pupils' tracking. Empirical evidence from France</b>	<b>23</b>
1.1 Introduction	25
1.2 Related literature	26
1.3 Empirical setting and Data	28
1.3.1 Description of the French education system	28
1.3.2 Data description	32
1.4 Teachers' grading	34
1.4.1 Descriptive statistics on students' test scores	34
1.4.2 Empirical evidence for grading differentials	36
1.5 Tracking	40
1.5.1 Descriptive statistics on tracking	40
1.5.2 Students' track choice independently of grading	42
1.5.3 Impact of grading differentials on the track choice	45
1.5.4 Teachers' track recommendation independently of grading	47
1.5.5 Direct measures of teachers' impact on tracking	51
1.6 Discussion	54
1.6.1 Interpreting the in-class score as the <i>perceived</i> academic ability	54
1.6.2 Interpreting the final exam score as the <i>true</i> academic ability	55
1.7 Conclusion	57
Appendices	58
Appendix A: defining students' SES	58
Appendix B: Additional tables	59



<b>2</b>	<b>Individual contribution in team contests</b>	<b>63</b>
2.1	Introduction . . . . .	65
2.2	Related literature . . . . .	69
2.3	Testing for neutrality in multiple pairwise battles . . . . .	71
2.3.1	Theoretical framework . . . . .	71
2.3.2	Empirical setting and data . . . . .	74
2.3.3	Testing for neutrality in multiple pairwise battles . . . . .	79
2.4	The role of individual contribution to team success . . . . .	84
2.4.1	Evidence for individual contribution . . . . .	84
2.4.2	Robustness checks and alternative explanations . . . . .	89
2.5	Conclusion . . . . .	97
	Appendices . . . . .	99
	Appendix A - theoretical predictions	
	. . . . .	99
	Appendix B - probit estimation . . . . .	110
<b>3</b>	<b>Estimating individual productivity in football</b>	<b>113</b>
3.1	Introduction . . . . .	115
3.2	Related literature . . . . .	118
3.2.1	The “box-score” approach . . . . .	119
3.2.2	The fixed-effect approach . . . . .	119
3.3	Data and estimation strategy . . . . .	121
3.3.1	Data . . . . .	121
3.3.2	The structural equation . . . . .	122
3.3.3	Estimation procedure . . . . .	124
3.4	Results . . . . .	129
3.4.1	Interpreting the coefficients . . . . .	129
3.4.2	The most productive players . . . . .	130
3.4.3	Comparing our measure of productivity to the market value of players . . . . .	134
3.4.4	A market test for discrimination . . . . .	136
3.5	Robustness . . . . .	139
3.5.1	In-sample predictions with simulations of seasons . . . . .	139

3.5.2	Out of sample predictions . . . . .	142
3.6	Extensions . . . . .	145
3.6.1	Introducing time-varying productivity . . . . .	145
3.6.2	Application of our estimation procedure to other team settings . . . . .	146
3.7	Conclusion . . . . .	147
Appendices . . . . .		148
Appendix A: Ordered probit . . . . .		148
Appendix B: Productivity ranking (top 100 - all positions) . . . . .		149



# Summary

## Three essays in applied economics

In this dissertation, I investigate several questions of interest in the fields of economics of education and economics of teams.

In the first chapter, based on initial work with Laurent Rossignol, we focus on tracking in education. Tracking refers to the fact that many countries' public school systems divide students by ability. This process can take different forms. In the United States and Canada, students are sorted into different classrooms within a school, whereas in most European countries, students are sorted in different schools with different curricula at the secondary level (Betts et al. (2011)). We provide evidence of tracking gaps in the French education system: tracking does not solely depend on academic performance but also on gender and socio-economic status. Our main contribution is to disentangle the impact of differences in students' aspirations from the impact of teachers' grading and track recommendations on these tracking gaps.

In the two following chapters, we leave the economics of education for the economics of teams. Teamwork has become more and more prominent in organizations (Hamilton, Nickerson & Owan (2003)). This raises some important economic questions. Why is forming teams economically desirable? What are the incentives associated to teamwork? How can individual productivity be assessed in teams, when only the collective output is observed? How does the team composition affect productivity? Are teams more productive if their members are homogeneous? In chapters 2 and 3 of this thesis, we use sports data to deal with some of these questions. As sports are settings where rules are relatively simple and data are often publicly available, they appear as a great laboratory to investigate the question of individual contribution in teams and to test economic theories of incentives and labor market behavior (Szymanski (2003)). The second chapter, joint with Antoine Chapsal, aims at understanding some of the incentives and psychological effects associated to teamwork, based on team squash data. We show that players value the fact of being responsible for the success of their team, which partly explains that team-based contests can yield higher effort than individual-based contests. The third chapter, built upon initial work with Rodrigo Lopez-Kolkovsky, aims at developing an estimation procedure to measure individual productivity in teams, based on European football data. We confront this measure to players' market value and provide evidence for racial

discrimination on the football market.

## **Chapter 1 - The impact of teachers on pupils' tracking. Empirical evidence from France**

My first chapter, based on a collaboration with Laurent Rossignol, focuses on tracking, which can be loosely defined as ability grouping in education. Our starting point is that tracking does not solely depend on students' academic performance. Indeed, a vast literature has shown the existence of tracking biases: individual characteristics such as race, gender and socio-economic status affect the track assignment, independently of academic ability (see Duru-Bellat, Jarousse & Mingat (1993) for an analysis of the French tracking system).

These differences in tracking can be partly explained by differences in students' preferences or aspirations, as emphasized by Guyon & Huillery (2016). However, as teachers take part in the track assignment procedure, differences in aspirations might not be the only driver of differences in tracking outcomes. In fact, teachers may have an indirect impact on tracking through their grading because grades are supposed to reflect students' academic potential and their chance of succeeding in the different tracks. Teachers may also exert a direct effect on tracking through their track recommendations if they are less prone to encourage some categories of students to follow the high ability track. Thus, assessing the role of teachers on grading and tracking is essential to better understand tracking gaps.

We use a very rich dataset on all ninth-graders from one of the 26 education districts in continental France (N=16864). Our data gather information on students' characteristics (SES, gender, nationality), academic performance, and track assignment procedure.

Our empirical setting appears to be ideal to study teachers' role on grading and tracking because it exhibits two important features. First, students take both in-class tests which are designed and corrected by their own teachers and a final exam which is common to the entire district and corrected in a blind manner. This allows us to understand teachers' grading practices. Second, students make an initial track choice to which teachers reply, which allows us to disentangle the effect of students' aspirations from the effect of teachers' recommendations on tracking.

We start by investigating teachers' grading by comparing students' in-class score to their final exam score. Our identifying assumption is that the in-class score reflects pupils' *observed/perceived* academic performance, whereas the final exam score reflects pupils' *true* academic ability. This assumption can be justified by several features of the setting: (i) As the final exam takes place at the very end of the academic year, only in-class grades are observed over the tracking procedure; (ii) In-class tests are corrected by pupils' own teachers and may be subject to grading bias, whereas the final exam is corrected in a blind manner, which makes it a purer measure of academic ability; (iii) The design of in-class tests is completely left at teachers' discretion so they are highly dependent on contextual specificity. On the contrary, the final exam is a standardized test that is common to all students in our sample; (iv) In-class tests are typically short and only cover the last lecture of the course, whereas the final exam covers all the material of grade 9 and better reflects long-term educational achievements.

We find that at equivalent score in the final exam, girls and high-SES pupils get an in-class score that is respectively 0.16 and 0.1 SD higher than the one of boys and low-SES pupils from the same class. Given our identifying assumption, we interpret these results as evidence for teachers' grading differentials.

Then, we focus on the tracking procedure. We provide evidence of strong differences in aspirations by SES and gender. High-SES pupils and girls are respectively 10.2 pp. and 3.4 pp. more likely to choose the high-ability track than low-SES pupils and boys from the same class with the same in-class score.

Integrating our analysis on grading differentials into our model of pupils' initial track choice allows us to identify the indirect effect of teachers' grading on pupils' track choice. Indeed, when we use the final exam score instead of the in-class score, the socio-economic and gender gaps in initial track choice increase to respectively 12.3 pp and 6.7 pp. This is evidence that teachers' grading indirectly affect tracking.

Turning to the analysis of teachers' track recommendations allows us to estimate the direct impact of teachers on tracking. Comparing the coefficients associated to SES and gender with the initial track choice and the teachers' track recommendation, at equivalent in-class score, allows us to identify whether teachers use their recommendations to reduce gaps in the initial track choice. We show that teachers' recommendations do not "correct" the gender gap in aspirations but slightly reduce the socio-economic gap in aspirations (the effect goes from 10.2 pp to 5.4 pp).

Plugging our analysis on grading in our model of teachers' recommendations allows us to measure directly the overall impact of teachers on tracking, i.e. both the indirect effect of their grading on tracking and the direct effect of their track recommendations. We show that, overall, teachers widens the initial gender gap in aspirations. This is coherent with our previous findings: girls get higher in-class grades than equally achieving boys and teachers do not use their track recommendations to lower the gender gap in initial choice. The total impact of teachers on tracking with regards to pupils' SES is more contrasted. On one hand, their grading disadvantages low-SES pupils which indirectly affect pupils' initial track choice and teachers' track recommendations. On the other hand, teachers use their track recommendations to slightly reduce the initial gap in aspirations between high-SES and low-SES pupils. Our results show that these two opposite effects cancel out and that globally, teachers do neither accentuate nor reduce the initial socio-economic gap in aspirations.

Overall, the main contribution of this chapter is to build a bridge between grading and tracking in education, which are usually treated as two separate questions. This allows us, not only to estimate teachers' grading differentials, but also their indirect impact on pupils' tracking.

## **Chapter 2 - Individual contribution in team contests**

The starting point of my second chapter, joint with Antoine Chapsal, is that despite the free-riding incentives that are inherent to teamwork (Alchian & Demsetz (1972)), organizing work in teams has become the linchpin in most organizations. Hence, the negative effect of free-riding incentives must be offset by positive effects associated to teamwork. Kandel & Lazear (1992) show in a theoretical paper that peer-pressure, guilt and shame are effects that mitigate the free-riding incentives. In this paper, we focus on another potential incentive that may arise in teams and counterbalance free-riding: the "individual contribution effect", according to which workers value the fact of being (at least partly) responsible for their team's success. Indeed, such an effect could explain why efforts may be higher in teams than in individual settings and partly explain why teamwork has become so prevalent.

As discussed in the last chapter of this dissertation, assessing individual contributions to the team output is very challenging. Therefore, we need to focus on a controlled team setting in which individual contribution is observable. We concentrate on multiple pairwise battles,



which refer to extremely common situations where players from two rival teams compete in individual battles (see Fu, Lu & Pan (2015) for a benchmark theoretical analysis of this kind of contests). As every player is responsible for his own individual battle, the assessment of individual contributions to the team output is straightforward.

Squash team championships exactly correspond to a best-of-three multiple pairwise battles setting. In a team squash contest, three distinctive pairs of players from two rival national teams are matched in three individual matches. Individual matches are played sequentially and a team needs two individual wins in order to win the contest. In this setting, the individual contribution is observed, since a player brings one point to his team if he wins and zero if he loses.

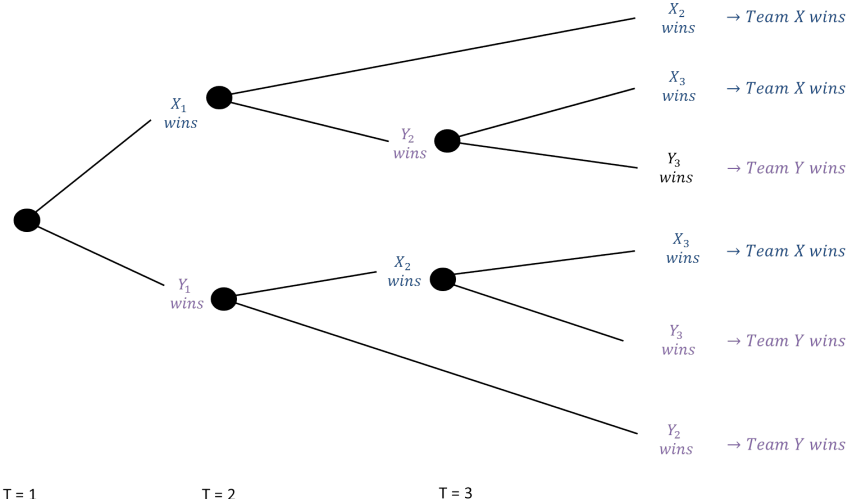


Figure 1: Squash team confrontations - a perfect example of contest with multiple pairwise battles

Furthermore, team squash contests exhibit an important random feature that makes them an ideal laboratory to study multiple-pairwise battles. In the beginning of the championship, every team has to rank its players by descending order of strength. Then, before each team contest, the order of individual games is randomly drawn from four possibilities: 1-2-3, meaning that players ranked first play the first game, players ranked second play the second game, and players ranked third play the third game, 1-3-2, 2-1-3 and 3-1-2. This ex-ante randomly-drawn order of play ensures that teams cannot manipulate the sequence of games to be played.

We collected from a specialized website data on World Team Championships, European Team Championships and Asian Team Championships, along with the monthly world rankings

of participating players based on their performance in individual tournaments.

According to the individual contribution effect, a player gets an extra payoff when he contributes to the success of his team (i.e. when he wins his individual match and his team wins the contest). In order to test for this effect in the squash data, we assess whether a player is more likely to win the second individual match when his teammate won the first match. The intuition behind this test is that in the second match, there is an asymmetry between the two players. The player in the leading team knows that winning his individual match implies that his team wins the contest and that he will get the “contribution reward”. On the other hand, the player in the lagging team will get the contribution reward if he wins only if his teammate also wins the third individual match. Thus, the two players do not face the same contribution incentives, which makes the player in the leading team more likely to win. To test this effect empirically, we start by focusing on a subsample of players who have very similar rankings. We find that the player in the leading team wins about 60% of the time, which confirms the individual contribution effect. In another specification, we use the whole sample and we include the ratio of rankings or rankings modalities to control for players’ relative ability. With these specifications, we find that the player in the leading team is more likely to win than the player in the lagging team by about 10 percentage points. These findings argue in favour of the individual contribution effect.

As an additional test, we check whether the expected outcome of the third individual match has an impact on the second individual match. Indeed, in the second match, if the player in the lagging team is sure that his teammate will lose the third match, the asymmetry of contribution incentives reaches its maximum as he will not get the contribution reward, whatever happens in his individual match. On the contrary, if he is sure that his teammate will win the third match, there is no asymmetry anymore as the second match becomes decisive for the two opposing players. Hence, according to the individual contribution effect, the probability of winning the second match increases with the expected probability of winning the third match. We provide empirical evidence showing that a player is more likely to win the second match when his teammate in the third match has a better ranking than his opponent.

In our last test, we focus on trivial third individual matches (i.e. third matches for which the winning team has already been determined over the two previous matches). One might argue that the individual contribution effect is also at play, to a lesser extent, in these battles. The player in the team that won the two first matches has more incentives to win than his opponent

because he can contribute to the success of his team. On the other hand, the player in the team that lost the two first matches, cannot participate to his team's success. We provide evidence that the player in the team that won the first two individual matches is more likely to win the third match than his opponent, controlling for the two players' relative ability. This confirms the individual contribution effect. As robustness checks, we rule out other possible mechanisms that could explain that winning the first match increases the probability of winning the second match: guilt-aversion, altruism, choking under pressure and psychological momentum.

Overall, we use team squash data to provide evidence of a dynamic linkage between individual matches in a multiple-pairwise battle setting. We provide evidence that this team dynamic is driven by an individual contribution effect: players value the fact of contributing to their team's success. This effect could mitigate free-riding incentives in teams, and be one of the explanations of the widespread use of teamwork in most organizations.

### **Chapter 3 - Estimating individual productivity in football**

My third chapter, based on initial work with with Rodrigo Lopez-Kolkovsky, aims at measuring football players' individual productivity. Our contribution is two-fold. First, from a methodological point of view, we develop an estimation procedure that could be applied to other team settings. Second, getting a reliable measure of individual productivity allows us to study the functioning of the labor market. Indeed, we can compare our measure of productivity to players' market value in order to test whether there is discrimination on the players' market. This approach is appealing for it allows us to test for discrimination directly at the player's level. We provide evidence of racial discrimination towards black players.

Football appears as an ideal laboratory to explore the question of assessing individual productivity in teams for three main reasons. First, it is a typical example of setting where only the collective production is directly observable. While the number of goals or points of the team is known, it is difficult to assess to what extent every player contributes to it. Second, rules are clearly defined, which makes it a controlled environment. Third, the recent explosion of available field data on football makes the analysis of the "beautiful game" possible.

Two approaches can be considered in order to assess players' individual productivity. The

first approach is based on players' individual statistics such as passes and shots. After identifying the variables that make a team successful, players' performance is assessed according to their participation to these variables (see for instance Berri & Schmidt (2010)). The second approach, on which our estimation procedure relies, is based on players' fixed-effects related to dummy variables indicating whether they are on the field. It is completely agnostic regarding what makes a player productive. The idea is to assess players' performance by comparing their team's performance when they are on the field and when they are not, controlling for their teammates and their opponents. It draws on the plus-minus statistic developed in ice-hockey in the US in the 1960s and on its adjusted versions (Rosenbaum (2004) and Ilardi & Barzilai (2008)). This approach is very low data-demanding as only the line-ups and the final scores are needed in the estimation. It also has the advantage of considering individual productivity in all its dimensions, including the ones that are not available in the data such as players' charisma.

We collected data on line-ups and final scores on four different websites in order to construct an extensive dataset of all the games opposing two teams from the five main European championships between 2007/2008 and 2014/2015.

The identification of players' fixed effects is possible thanks to the variability in the line-ups over time, due to coaching strategies, transfers and injuries. However, the estimation is challenging as the number of coefficients to be estimated is very large: in our setting the ratio between the number of observations and the number of variables is only 8. Furthermore, there is a strong colinearity issue since players often play with the same teammates. These two characteristics of the setting imply that a standard estimation of players' fixed effects will suffer from overfitting: if players A and B are always together on the field except during one game in which only player A is in the line-up and their team is very successful, a standard estimation will lead to a much higher coefficient for player A than for player B. This is not desirable, as the performance differential between the two players will be based on one game only. In other words, standard estimation techniques will lead to a very large variance of coefficients and poor out of sample predictions. Therefore, we need to adjust the estimation procedure, so that it yields different coefficients to two teammates if and only if there are enough games in which the two players are not together on the field.

Ridge is a penalization method that is particularly well-suited to do that. The intuition behind the Ridge is that the estimation trades off between the goodness of fit (just as in a standard estimation) and the variance of the coefficients, by adding a penalization term corresponding to

the sum of the squared estimated coefficients. Hence, it will yield large coefficients to players only if it has enough observations and variations to be sure that these players are high-achieving players. The Ridge solves the overfitting problem. We choose the optimal degree of penalization  $\lambda$  by 10-fold cross-validation.

We obtain results that are coherent with football specialists' point of view. Lionel Messi and Cristiano Ronaldo, who are largely recognized as the best players over the period, top our ranking of players' fixed effects. All other players in our top 30 are world class players. This is very satisfactory given that the model is completely agnostic regarding what makes a player productive. As a robustness check, we use several in-sample and out-of sample tests at the team level. The intuition behind these tests is that players are often transferred from one club to another. Therefore, if the aggregate level of teams is well predicted, then the level of individual players is also well predicted because compensation of players' coefficients is unlikely.

Table 1: Overall ranking of players : top 10

Ranking	Player	Total	Offensive	Defensive
1	Lionel Messi	.351	.351	
2	Cristiano Ronaldo	.333	.333	
3	Cesc Fabregas	.302	.243	.059
4	Frank Lampard	.274	.205	.069
5	Yaya Toure	.255	.138	.117
6	Karim Benzema	.253	.253	
7	Bastian Schweinsteiger	.246	.171	.075
8	Xavi	.241	.175	.066
9	Arturo Vidal	.238	.128	.11
10	Franck Ribery	.237	.237	

Seasons 2007/2008 to 2014/2015.

Our measure of players' performance suffers from several limitations. First, the Ridge penalization implies that the estimation only provides points estimates, and not the standard errors of players' fixed effects. This prevents us from testing whether two players have a significantly different productivity. Second, the coefficients are difficult to interpret. To remedy that, we convert them into marginal effects, but this transformation ignores the bias introduced by the Ridge. Third, in order to get enough variability in the data, we need to impose restrictions on the coefficient dynamics: players' productivity is considered fixed over the entire time-span;

players' productivities do not depend on their teammates; defenders do not have offensive fixed effects and forwards do not have defensive fixed effects.

Despite these limitations, the fact that we obtain results that make so much sense shows that combining a fixed-effect approach with a Ridge penalization is an estimation strategy that is very well-suited to the estimation of individual productivity in teams. The fixed-effect approach allows us to consider productivity in all its dimensions (including the unobservable ones) while the Ridge allows us to overcome the problem of isolating individual productivity when players often play with the same teammates. Thus, our estimation procedure could be adapted to other team settings where assessing individual productivity is important. The only requirements are that workers stay long enough in the firm (number of observations) and that shifts in the teams' compositions are frequent (variability).

Once our productivity metrics obtained, we use it in order to test for racial discrimination on the football market. We compare our measure of productivity to players' market value according to the German website Transfermarkt. The two measures are strongly correlated, which confirms that our estimation procedure leads to results that are globally in line with the market. However, the fit is not perfect and we can identify undervalued and overvalued players. We compare the share of black players in the two groups and we show that they are over-represented among undervalued players. Our approach complements the existing literature on discrimination in sports, which is based on empirical tests for discrimination at the team level (see in particular Szymanski (2000)).

## **Appendix - applied statistics in cardiology**

During my PhD, I also participated in a research project with cardiologists from the CHU Nancy<sup>1</sup> and the French football league<sup>2</sup>. Our aim was to establish the cardiac profile of highly trained football players.

I was in charge of the data management and the statistical analysis. I coauthored a paper about the electrocardiographic profile of football players which was published in the *Archives of*

---

<sup>1</sup>Service de cardiologie, institut Lorrain du cœur et des vaisseaux Louis-Mathieu, centre hospitalier universitaire de Nancy, 4, rue du Morvan, 54500 Vandœuvre-lès-Nancy, France

<sup>2</sup>Ligue de Football Professionnel, 6 rue Léo Delibes, 75116 Paris

*Cardiovascular Diseases*.<sup>3</sup> We are currently working on other articles that should be published in 2019.

---

<sup>3</sup>Huttin, O., Selton-Suty, C., Venner, C., Vilain, J. B., Rochecongar, P., & Aliot, E. (2017). Electrocardiographic patterns and long-term training-induced time changes in 2484 elite football players. *Archives of cardiovascular diseases*.

## Chapter 1

# The impact of teachers on pupils' tracking. Empirical evidence from France

JEAN-BAPTISTE VILAIN & LAURENT ROSSIGNOL



## **Abstract**

In education, grading and tracking are two interconnected questions as grades are supposed to reflect students' academic ability and their chance of succeeding in the different tracks. Using a very rich data set on French ninth graders in which we have information on (i) both blind and non-blind test scores and (ii) both the initial track choice of students and the teacher conference's track recommendation, we show that teachers play an important role in pupils' track assignment. First, teachers' grading of non-blind tests depend on students' gender and socio-economic status, which indirectly affect tracking. Second, teachers do not use their track recommendations to erase the socio-economic and gender gaps in the initial track choice.

**JEL Classification** I24, I21, J16, J15.

**Keywords** Education, Tracking, Grading, Teachers, Discrimination.

## 1.1 Introduction

Tracking refers to “the tendency in many countries’ public school systems to divide students by ability in some way. Students might be sorted into different classrooms within a school, or sorted into different schools” (Betts et al. (2011)). Within-school tracking is common in the US, the UK, Canada and Japan while most European countries track students into different schools with either an academic or a vocational curriculum.

In a purely meritocratic world where students have homogeneous preferences, tracking should only be affected by students’ academic ability. However, an extensive literature has shown the existence of *tracking biases*: some individual characteristics such as gender, race or socio-economic status (henceforth SES) affect tracking, independently of academic performance. Most of the literature focuses on students’ aspirations to explain these tracking biases. For example, Hoxby & Avery (2013) show that among top US students, low-income students are less likely to apply to top universities than high-income students, despite the fact that top universities would not cost them more money. Guyon & Huillery (2016) use data on French 9th graders to show that at equivalent academic potential, low-SES students have lower aspirations than high-SES students.<sup>1</sup>

However, students’ track choice may not be the only determinant of tracking biases, as other actors take part in the track assignment procedure. In particular, teachers may play an active role on tracking through two different channels. First, teachers may be more severe when they grade some categories of students. Tracking then heavily depends on students’ grades as they are the main source of information regarding pupils’ ability. Second, teachers’ tracking recommendations may depend on students’ characteristics. At equivalent academic potential, teachers may be less prone to recommend some categories of students the “high-ability” track.

This paper aims at understanding some of the mechanisms that cause tracking biases. We disentangle the impact of students’ preferences (or aspirations) from the impact of schools and teachers on tracking. To do so, we use a very rich dataset on French students in grade 9 in which we have information on (i) both blind and non-blind test scores, which allows us to understand teachers’ grading practices and (ii) both the initial track choice of students and the teacher conference’s opinion on this track choice, which allows us to distinguish the impact of teachers’

---

<sup>1</sup>The content of their dataset is similar to ours but they focus on different French education districts and their sample only includes part of the students from the districts under consideration.

recommendations from the impact of students' choices on the track assignment.

The main results are as follows:

1. Evidence of grading differentials: at equivalent score in the final standardized blind exam, girls and high-SES pupils get higher in-class grades than boys and low-SES pupils from the same classroom.
2. Independently of grading differentials, aspirations are gender and socially determined: girls and high-SES apply much more often to the high-ability track than equally-achieving boys and low-SES students. This result is consistent with the existing literature on tracking and aspirations (see Guyon & Huillery (2016)).
3. Independently of grading differentials, teachers' recommendations do not correct the gender gap in aspirations, but they slightly reduce the socio-economic gap in aspirations.
4. Teacher's grading aggravates tracking gaps. As in-class grades are the only available source of information regarding students' performance, the "true" academic ability of boys and low-SES pupils is under-estimated. This affects tracking through both students' initial track choice (low-SES and boys under-estimate their ability, which makes them less likely to apply to the high-ability track) and teachers' track recommendation (teachers also underestimate the academic ability of these students, which makes them less likely to recommend them the high-ability track).

The remainder of this paper is organized as follows. Section 1.2 presents the related literature on grading and tracking. Section 1.3 presents the empirical setting under scrutiny and the available field data. Section 1.4 presents our key identifying assumption and provides evidence of grading differentials. In section 1.5, we disentangle the impact of students' choice from the impact of teachers on the track assignment. We discuss the limits of our identifying assumption in section 1.6 and we conclude in section 1.7.

## 1.2 Related literature

Teacher expectations might depend on students' characteristics. Tenenbaum & Ruck (2007) show that teachers in the US have different expectations regarding the performance of pupils according to their ethnic background. Darley & Gross (1983) find experimental evidence that

teachers have lower expectations for low-SES students. If teachers have low expectations concerning some groups of students, they might look more carefully for errors or weaknesses when evaluating their work. This expectation confirmation bias may in turn affect the way teachers evaluate students' academic potential and motivates us to study the role of teachers on grading and tracking.

**Grading** From an empirical point of view, testing whether discrimination exists is difficult for disadvantaged groups come from disadvantaged backgrounds with characteristics that are correlated with poor academic performance. Therefore, it is hard to disentangle the discrimination effect from the characteristics effect. Studies on discrimination in grading have used two main identification strategies to solve this issue.

The first approach is experimental: teachers are asked to evaluate academic tests for which the researcher has manipulated the characteristics of student to whom the work is attributed. For example, Hanna & Linden (2012) test for the existence of grading discrimination in India by randomly assigning child characteristics (age, gender and caste) to the front page of the exams. They find evidence of grading discrimination against lower caste children. Sprietsma (2013) randomly assigns typical German or Turkish names to identical sets of essays to test whether teachers' expectations are different for migrant children in Germany. She finds that essays bearing Turkish names receive significantly worse grades. Van Ewijk (2011) conducted a similar experiment in the Netherlands but did not find any effect of names on grades.

The second approach is empirical and based on observational data: it compares the academic performance in class tests ("non-blind" grading) to the academic performance in anonymous tests ("blind" grading). Based on Israeli data, Lavy (2008) finds that male students face grading discrimination in humanities and sciences. Breda & Ly (2015) use a similar identification strategy to show that evaluation is biased in favor of females in more male-dominated subjects (e.g. math) and in favor of males in more female dominated subjects such as literature and biology. On a much broader sample, Breda & Hillion (2016) show that evaluation is biased in favor of women in all subjects and that the magnitude of this bias increases with the degree of subject's male-domination. We adopt a similar approach to these studies in order to understand teachers' grading. Then, we integrate this analysis on grading in a tracking model. This allows us to identify, not only grading differentials, but also their indirect effect on students' tracking.

**Tracking** At equivalent academic potential, teachers’ track recommendations might not be the same with every student. Teachers might be less prone to encourage some categories of students to engage in the high-ability track. Elhoweris, Mutua, Alsheikh & Holloway (2005) show that the children’s ethnicity affect US teachers’ referral and placement decisions in gifted and talented programs. Glock, Krolak-Schwerdt, Klapproth & Böhmer (2013) also investigate the question of ethnic tracking biases and find evidence of less favorable teacher judgments of students with immigrant backgrounds than of students without immigrant background. Frey (2002) find that SES affects special education teachers’ placement recommendations: children from low SES backgrounds are more likely to be referred for restrictive placements.

These studies analyze teachers’ tracking recommendations independently from pupils’ aspirations and teachers’ grading practices. Our empirical setting allows us to take into account these two factors. First, as we observe both pupils’ initial track choice and teachers’ track recommendation, we can disentangle the effect of pupils’ aspirations from the effect of teachers’ recommendation on tracking. Second, integrating our analysis on grading into our model of tracking allows us to disentangle the direct influence of teachers on tracking (through their track recommendations) from their indirect influence (through their grading practices).

Overall, our main contribution in regards to the existing literature is to build a bridge between grading and tracking in education, which are usually treated as two separate questions.

## 1.3 Empirical setting and Data

This section describes the grading and tracking system in France along with the available data and descriptive statistics.<sup>2</sup>

### 1.3.1 Description of the French education system

**Tracking** In France, students follow the same curriculum across schools from kindergarten to the end of junior high school. Junior high school runs from grades 6 to 9. After junior high school, pupils are tracked either in a vocational track or in the academic track. Most of senior high schools are specialized in either academic training (*Lycée général et technologique*) or vocational training (*Lycée professionnel*) while some of them offer both (*Lycée polyvalent*). In the end of a vocational track, students get a professional degree allowing them to find a job

---

<sup>2</sup>Note that the French education system has changed very recently, in particular in regards grading. The description provided in this paper applies to the academic year for which we have data (2013/2014).

with no further education while students enrolled in the academic track get a general degree and are expected to get higher education. Students enrolled in the academic track tend to have much better academic performances in junior high school than students enrolled in a vocational track. In our sample, the average final exam score at the end of grade 9 is 11.1/20 (0.5 SD)<sup>3</sup> for those in the academic track and 6.6/20 (-0.8 SD) for those in a vocational track.

The track assignment procedure is a process in which pupils, families and schools are involved. In January-February, pupils and their families indicate their track choice among four possibilities: the academic track (*2nde générale et technologique*), a 3-year vocational track (*Brevet d'études professionnelles*), a 2-year vocational track (*Certificat d'aptitude professionnelle*) or grade repetition (which is quite common in France, contrary to other countries<sup>4</sup>). In March, the teacher conference<sup>5</sup> gives its opinion on this initial choice, either encouraging it or suggesting another track. In May, pupils and their families make their final track choice, which is then validated or invalidated by the teacher conference's decision in June. An “academic track decision” grants access to both the academic track and any vocational track whereas a “vocational track decision” only grants access to a vocational track.<sup>6</sup> This asymmetry is justified by the implicit norm that a student who is expected to succeed in the academic track is also expected to succeed in the vocational track, while the reverse is not true. Figure 1.1 summarizes the tracking procedure over grade 9 in France.

---

<sup>3</sup>As usual in the literature, we standardize students' test scores to distributions with zero mean and a unit standard deviation. Accordingly, a final exam score of 11.1/20 corresponds to a standardized final exam score of 0.5 SD. This standardization makes results easier to interpret across different education systems.

<sup>4</sup>See Goux & Maurin (2007)

<sup>5</sup>The teacher conference gathers the provost, teachers of every subject, two class representatives and two parent representatives. It takes place three times a year for every class and it aims at discussing students' quarterly academic performance.

<sup>6</sup>Note that the family can request an appeal committee in case of disagreement.

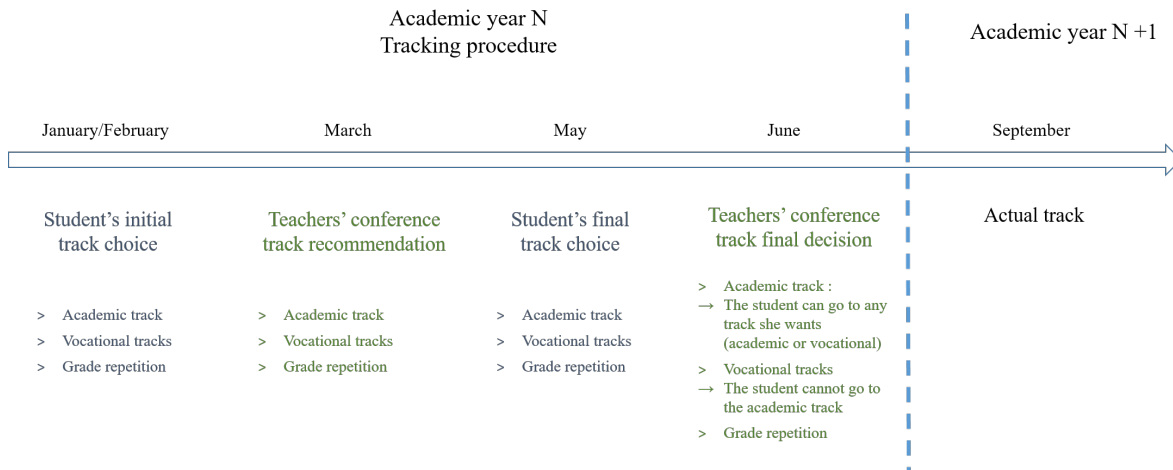


Figure 1.1: Summary of the tracking procedure

The sequential nature of the process between students' choices and teacher conference's responses is a key feature of our identification strategy because it allows us to distinguish the impact of students' preferences and expectations from the impact of teachers and schools on tracking. Though, an important limitation of our study regarding the sequence of the tracking procedure is that we only have data on pupils' average in-class score over the entire academic year, and not disaggregated data by academic quarters. Hence, we cannot consider pupils' progress over the year and we implicitly assume that academic progress over grade 9 is independent from gender and SES.

**Grading** In grade 9, pupils take two different types of tests. From September to June, they take many in-class tests which are corrected by their own teachers in a non-blind manner (*contrôle continu*). In-class tests are taken in all the subjects of grade 9 curriculum: Mathematics, French, History/Geography, Physics/Chemistry, Biology/Geology, Foreign language 1, Foreign language 2, Physical education, Information and Technology, Art and Music. Every subject gets the same weight in the *contrôle continu*. The design of in-class tests is completely left at teachers' discretion. Teachers decide on the number of in-class tests to be taken, their format and their difficulty. They are also free to adjust the severity of their grading. The fact that teachers can organize in-class tests as they want implies that in-class scores of students from different classes are difficult to compare. Most of the time, in-class tests are short and only cover the last lecture of the course.

In the end of June, students take a final exam which is anonymously and externally graded

in July (*contrôle final*). This exam covers the three main subjects, *i.e.* the three subjects with the highest course load: Mathematics, French and History/Geography. The three subjects get the same weight in the *contrôle final*. The final exam is the same for all ninth-graders, which guarantees the comparability of the final exam score from one class to the other and the absence of contextual effects (e.g. easier tests in classes with weaker students). The final exam covers all the material of grade 9 and is supposed to reflect the overall academic achievement of pupils over junior high-school.

Our identification strategy to estimate grading differentials relies on the comparison of students' score in the in-class tests to their score in the final exam. Of course, we need to compare students' academic performance in subjects that are evaluated both in the final exam and in class. Therefore, in the remainder of the paper, by *in-class test score* we mean the average test score in class in Mathematics, French and History/Geography. In order to ensure that our results are not driven by heterogeneity in performances in minor subjects, we compare the relative performance of students in the three major subjects and in minor subjects by gender and SES. Table 1.1 represents the average ratio of the in-class score in minor subjects on the in-class score in the three major subjects by SES and gender. Low-SES pupils and boys perform relatively better in minor subjects than high-SES students and girls, so the tracking gaps are not likely to be driven by unobserved differences in performances in minor subjects.

Table 1.1: Average ratio of in-class scores

	<i>Average in-class score in minor subjects</i> <i>Average in-class score in the three major subjects</i>
<i>Students' characteristics</i>	
High SES	1.11
Intermediate SES	1.18
Low SES	1.21
Girl	1.17
Boy	1.2

Students' global in-class score is combined with their final exam score to determine whether they obtain the junior high school degree (*Diplôme national du brevet*).<sup>7</sup> Note that the final exam takes place after the track assignment procedure. Hence, getting the junior high school degree is not a necessary condition to enroll in the different tracks and grades obtained in the

<sup>7</sup>In our sample (2013/2014), the weight of in-class tests in the overall grading is 60%.



final exam are not observed during the track assignment procedure. Pupils and schools can only use in-class grades to assess pupils' academic potential and their likelihood of succeeding in the different tracks.

### 1.3.2 Data description

Our sample includes data on all ninth-graders in public schools from the district of Amiens, one of the 26 education districts in continental France, over the academic year 2013/2014. As we have data on all the students from the district, our sample is large and we do not face any selection issue regarding junior high-schools in the sample.

In order to construct our dataset, we merged information from different sources:<sup>8</sup>

- Data from the Amiens district (*Base élève académique*): pupils' date of birth, gender, nationality, SES, place of residence, along with information on the junior high-school and the class in which they are enrolled<sup>9</sup>.
- Data on the tracking procedure (*Base INTORI*): pupils' initial track choice, teacher conference's recommendation, pupils' final choice and teacher conference's final decision.
- Data on grades and actual track enrollment the next year (*Base Affelnet*): average grades obtained in the in-class tests over the academic year and final exam score in the different subjects.
- Data on junior high schools (*Base APAE*): address, urban/rural school, ZEP status (zones of educational priority), ZUS status (sensitive urban zone).

Overall our sample provides information on pupils' characteristics, academic performance and tracking for a population of 16864 students.<sup>10</sup> Table 1.2 presents a few summary statistics and compares our sample to the entire population of the 596053 ninth-graders in France in 2013/2014.<sup>11</sup> The comparison calls for the following remarks. Only 1.9% of pupils do not have the French nationality in our sample, which is twice lower than in the entire population. We

---

<sup>8</sup>The merge between datasets was made possible thanks to students' unique identifier.

<sup>9</sup>Having data on students' classes is crucial for it enables us to control for neighborhood, teacher and peer effects.

<sup>10</sup>We kept in our sample all initial observations for which we had information on gender, nationality, SES, class and the tracking process.

<sup>11</sup>Aggregate data on French ninth-graders and on French junior high-schools were provided to us by the DEPP (*Direction de l'évaluation, de la prospective et de la performance*).

define pupils' SES according to the *INSEE* classification of the two parents (see the precise description in the Appendix). When at least one parent is high-SES, the pupil is considered as high-SES. When none of the parents is high-SES and at least one of them is intermediate SES, the kid is considered as intermediate SES. In other cases, the pupil is considered low-SES. The proportion of high-SES students is 7 percentage points lower in our sample than in the French population.

Even though students' performance is slightly lower in our sample than in the entire population, we can identify very similar patterns. In-class test scores are lower in the three main subjects than in other subjects and the final exam score is much lower than the in-class score.

Junior high-schools in our sample are very representative of junior high-schools in France, as the proportion of schools labeled Zone of educational priority and the proportion of schools located in very disadvantaged urban neighborhood are almost the same (respectively 22% against 20.4% and 7.5% against 7.8%).

Table 1.2: Summary statistics and representativeness of the sample

Variable	Sample mean	Population mean
<i>Students' characteristics</i>		
Boy	0.486	0.497
Girl	0.514	0.503
French	0.981	0.962
Foreign	0.019	0.038
High SES	0.186	0.256
Intermediate SES	0.286	0.258
Low SES	0.528	0.486
N	16864	596053
<i>Academic performance</i>		
In-class tests (all topics)	12.1/20	13/20
In-class tests (Mathematics/French/History-Geography)	11.1/20	11.4/20
Final exam (Mathematics/French/History-Geography)	9.4/20	9.8/20
N	16864	596053
<i>Junior High-Schools' characteristics</i>		
Zone of educational priority ( <i>ZEP</i> )	0.220	0.204
Not in a zone of educational priority	0.780	0.796
Sensitive urban zone ( <i>ZUS</i> )	0.075	0.078
Not in a sensitive urban zone	0.925	0.922
N	173	5273

## 1.4 Teachers' grading

In-class grades have an effect on tracking because they are used by pupils, parents and teachers to evaluate the pupils' chances of succeeding in the different tracks. Indeed, only in-class grades are observed during the tracking procedure since the final exam takes place after the track assignment. Teachers may have an indirect impact on students' tracking through their grading of in-class tests. Studying whether teachers' grading depends on students' characteristics is thus essential to understand the role of teachers in tracking.

### 1.4.1 Descriptive statistics on students' test scores

**Academic performance** We standardize in-class test scores and the final exam score to distributions with zero mean and a unit standard deviation, so as to make the statistics easier to interpret. This procedure is applied within subjects.

Figure 1.2 displays the distribution of standardized test scores in the final exam and in class by SES. The Kolmogorov-Smirnov test shows that the three distributions are significantly different at the 1% level for both the final exam and in-class tests. On average, high-SES students outperform low-SES students by 0.88 SD in the final exam and 0.80 SD in class.

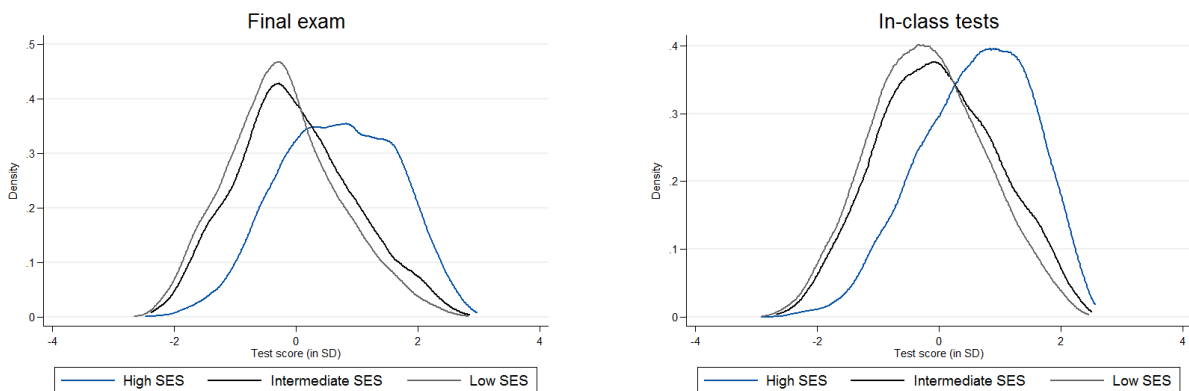


Figure 1.2: Distribution of academic performance by SES

Academic performance also depends on gender. Figure 1.3 displays the distributions of standardized test scores for boys and girls. The Kolmogorov-Smirnov test shows that the two distributions are significantly different at the 1% level for both the final exam and in-class tests. Though, the difference is much larger in in-class tests than in the final exam (the D value of the KS test is 0.085 in class against 0.036 in the final exam). On average girls outperform boys by

0.07 SD in the final exam and 0.23 SD in class.

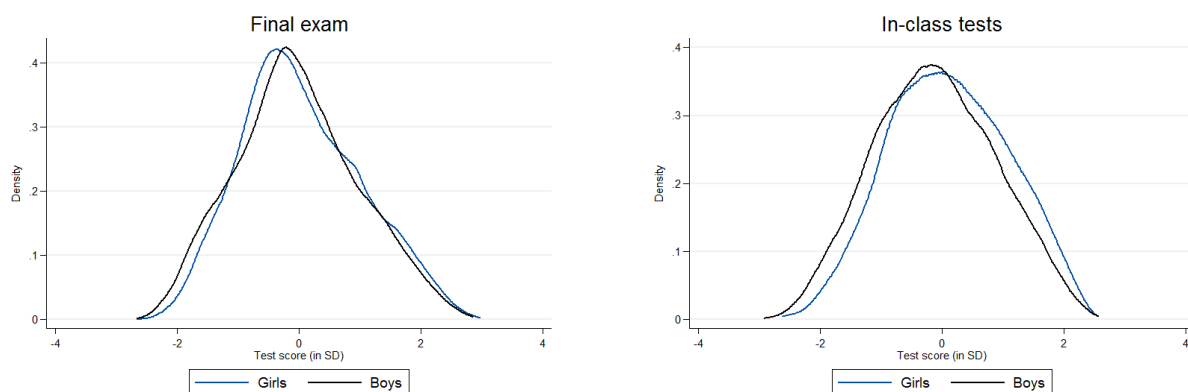


Figure 1.3: Academic performance by Gender

**The need for class fixed effects** As the design of in-class tests is left at teachers' discretion, the in-class score might be difficult to compare for students who are in different classes. As an illustration, figure 1.4 represents the final exam score and the in-class score of pupils in two different classes from the sample. Pupils from class B clearly outperform pupils from class A in the final exam but have similar grades in class, so we can conjecture that teachers in class B are much more demanding than teachers in class A. This example emphasizes the need to include class fixed-effects in our analysis of grading to account for such class specificity.

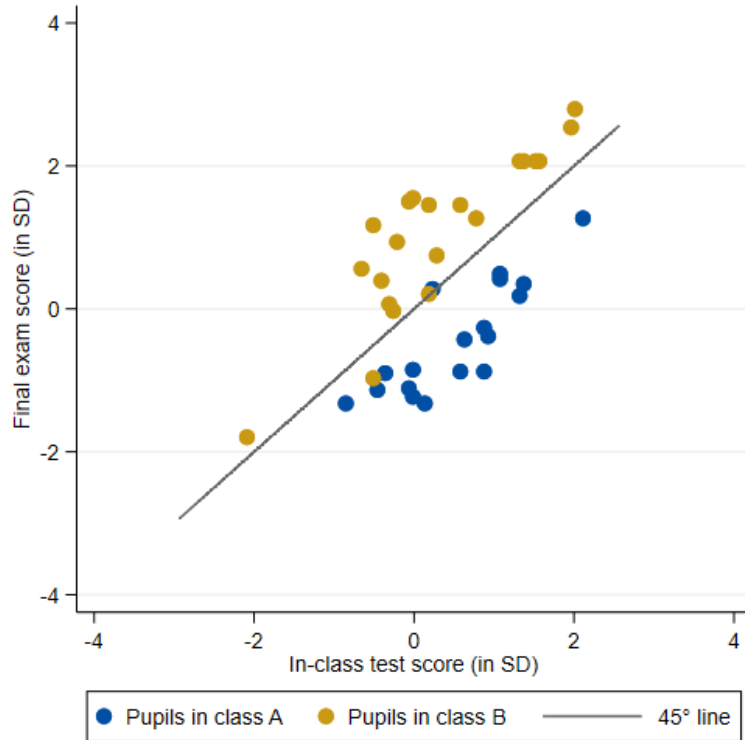


Figure 1.4: Example of two classes with different grading practices

### 1.4.2 Empirical evidence for grading differentials

As teachers are completely free in the design and organization of in-class tests, pupils' *observed* academic performance may not exactly correspond to their *true* academic potential.

**Model** We assume that the *observed/perceived* academic performance of student  $i$  in class  $j$  is common to the student and her teachers<sup>12</sup> and we denote it by  $\widetilde{\theta}_{ij}$ .

$\widetilde{\theta}_{ij}$  is of course related to the “*true*” academic ability of student  $i$ , that we denote  $\theta_i$  and that is unobserved during the academic year.  $\widetilde{\theta}_{ij}$  may also be affected by student  $i$ 's individual characteristics, in particular gender, SES and nationality. Indeed, students with different characteristics may be graded differently by teachers.  $\widetilde{\theta}_{ij}$  is also strongly impacted by the contextual effects associated to class  $j$ , as the perceived academic ability of students is in large part relative to the level of other students in the classroom. Finally,  $\widetilde{\theta}_{ij}$  depends on a random shock  $\epsilon_{ij}$  that

<sup>12</sup>We discuss this assumption further in section 1.6.

follows a standard normal distribution.

$$\begin{aligned} \widetilde{\theta}_{ij} = & \alpha_0 + \alpha_1\theta_i + \alpha_{girl}Girl_i + \alpha_{intSES}Intermediate\ SES_i + \alpha_{lowSES}Low\ SES_i \\ & + \alpha_{foreign}Foreign_i + \alpha_{FE}FE_j + \epsilon_{ij} \quad (1.1) \end{aligned}$$

**Identifying assumption** Given the respective characteristics of the in-class tests and of the final exam presented in section 1.3, we assume that the in-class score reflects the *perceived* academic ability  $\widetilde{\theta}_{ij}$  while the final exam score reflects the *true* academic ability  $\theta_i$ . This identifying assumption seems relevant for several reasons.

First, pupils and teachers only observe scores obtained in in-class tests. As the final exam is taken at the very end of the academic year after the track assignment, the final exam score is unobserved and cannot be used as a basis for perceptions of academic ability. The in-class score is the only available source of information about students’ academic potential.

Second, in-class tests are not blind, so they are subject to potential grading biases. On the contrary, the final exam is corrected in a blind and anonymous way, which makes it a “purer” measure of the *true* academic ability.

Third, the design and the grading severity of in-class tests strongly vary from one class to another, whereas the final exam is the same for every student. Therefore, the final exam score is a good measure of the *true* academic ability because it is independent of contextual specificity. Fourth, the scope of material covered in in-class tests is usually very limited. Most of the time, during in-class tests students are evaluated on content from the very last weeks of the course. Hence, the in-class score, which corresponds to the average performance in these small-scope tests, may not reflect long-term educational achievements. In contrast, the different exercises in the final exam are supposed to cover most of the topics of the program. Hence, performance in the final exam better reflects long-term achievements and the ability to mobilize a wide knowledge. In that sense, it better approximates *true* academic ability.

Thus, in the remainder of our analysis, we consider that the final exam score is a measure of the *true* academic ability  $\theta_i$  and that the in-class score is a measure of the *perceived* academic ability  $\widetilde{\theta}_{ij}$ . This assumption allows us to disentangle grading effects from tracking effects and is a key ingredient of our empirical strategy. We discuss and question this assumption further in section 1.6.

**Estimation** In order to estimate the  $\alpha$  parameters from equation 1.1, we regress the in-class score on the final exam score<sup>13</sup>, on student’s main characteristics (SES, gender, nationality) and on class fixed-effects to control for neighborhood, teacher and peer effects. As unobservables of pupils belonging to the same classroom might be correlated, we cluster standard errors at the class level.

**Results** Results of the OLS estimation are reported in the first column of table 1.3.

Table 1.3: Grading differentials

<i>Dep. var:</i>	In-class score (in SD)	
Final exam score (in SD)	0.907*** (0.004)	
Final exam score (in pctiles)		Yes
Girl	0.163*** (0.008)	0.162*** (0.008)
Intermediate SES	-0.086*** (0.011)	-0.088*** (0.011)
Low SES	-0.097*** (0.010)	-0.099*** (0.010)
Foreign	0.060* (0.030)	0.061* (0.030)
Constant	-0.003 (0.009)	-1.906*** (0.038)
Class fixed-effects	Yes	Yes
Observations	16864	16864
$R^2$	0.82	0.82

Standard errors are clustered at the class level and are reported in parentheses

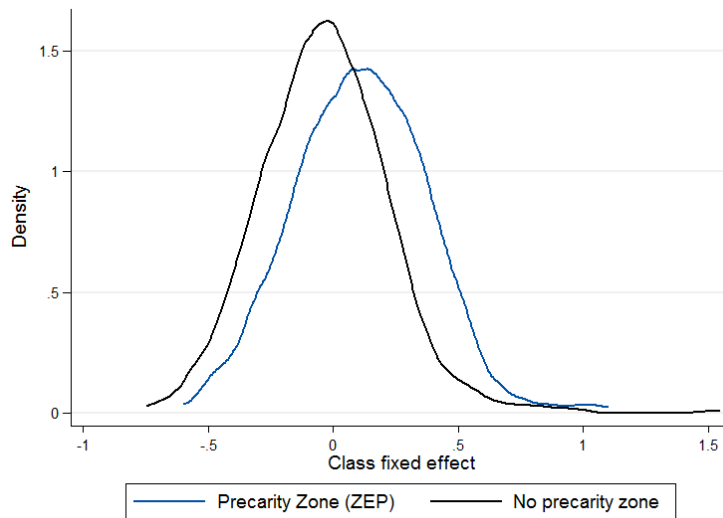
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Student’s SES and gender are significant at the 0.1% level. A girl gets an in-class score that is 0.163 SD higher than the one obtained by a boy from the same class who gets the same score in the final exam. Low-SES and intermediate SES students get lower in-class grades than high-SES students (respectively -0.097 SD and -0.086 SD). Foreign students get higher in-class grades than their equally achieving French classmates (+0.06 SD). In specification (2), we replace the

<sup>13</sup>As mentioned previously, the in-class score and the final exam score refer to the average grades obtained in the three main subjects (Mathematics, French, History/Geography).

continuous final exam score by dummies corresponding to the final score's percentile. This does not change the significance of the results and the magnitude of the effects remain in the same range, which shows that our results are not sensitive to the functional form used for the final exam score.

The analysis of class fixed-effects confirms that teachers' grading severity greatly varies from one class to another. Figure 1.5 plots the estimated in-class grades' fixed-effects from specification (1), distinguishing classes located in disadvantaged schools from others. Class fixed-effects are very strong in magnitude as there is a span of about 2 SD between the least severe and the most severe classes. This shows that teachers' in-class grading is far from being uniform in all school environments and justifies the need for a within-class identification strategy. Class fixed-effects tend to be much higher in disadvantaged neighborhood, which suggests that teachers' grading is less strict in classes where there are a lot of low-SES students.



Note: demeaned class-fixed effects estimated in specification (1) from table 1.3.

Figure 1.5: Distribution of estimated class fixed-effects in grading

As a robustness check, we run the same estimations subject by subject (instead of considering the average performance in the three subjects). The coefficients associated to gender and SES remain significant at the 0.1% level for the three subjects (see table 1.11 in the Appendix).

Overall, our results show that (i) teachers' grading greatly depends on the schools' context (between-class effect) and (ii) girls, high-SES students and foreign students get higher in-class grades than their classmates with the same final exam score (within-class effect). These grading



differentials might in turn affect tracking.

## 1.5 Tracking

As described in section 1.3, the tracking procedure is a process in which pupils' and their families make track choices that are validated or invalidated by the teacher conference. This institutional feature allows us to disentangle pupils' preferences and expectations from the influence of teachers and schools on tracking. Furthermore, by combining our analysis of tracking with our results on grading, we can identify the indirect effect of teachers' grading on tracking.

### 1.5.1 Descriptive statistics on tracking

**Tracking** The track assignment heavily depends on students' SES. In our sample, 87% of high SES ninth-graders are recommended the academic track the next year, against only 56.7% of low SES students. Girls are also more often recommended the academic track than boys (69% against 59.5%).

<b>Share of students who are recommended the academic track after grade 9</b>	
High SES	87%
Intermediate SES	64%
Low SES	56.7%
Girls	69%
Boys	59.5%
Total	64.4%

Table 1.4: Tracking by SES and Gender

These tracking differences may be partly driven by differences in academic performance, as high-SES students and girls tend to get higher test scores (see figures 1.2 and 1.3). However, even when we control for academic potential, the track assignment varies with pupils' SES and gender. In figure 1.6, we divide students in four groups based on their performance in the final exam (Group 1: top 25%, Group 2: 25-50%, Group 3: 50-75%, Group 4: bottom 25%). For each group, we report the share of students who are recommended the academic track the next year by SES. Among high-achieving students (top 25%), the vast majority of students are recommended the academic track, independently of their SES. When academic performance decreases, socioeconomic differences in tracking increase: in groups 2, 3 and 4, the share of high-SES students who are recommended the academic track is respectively 12 percentage points (pp.), 12 pp. and

16 pp. higher than the one of low-SES students.

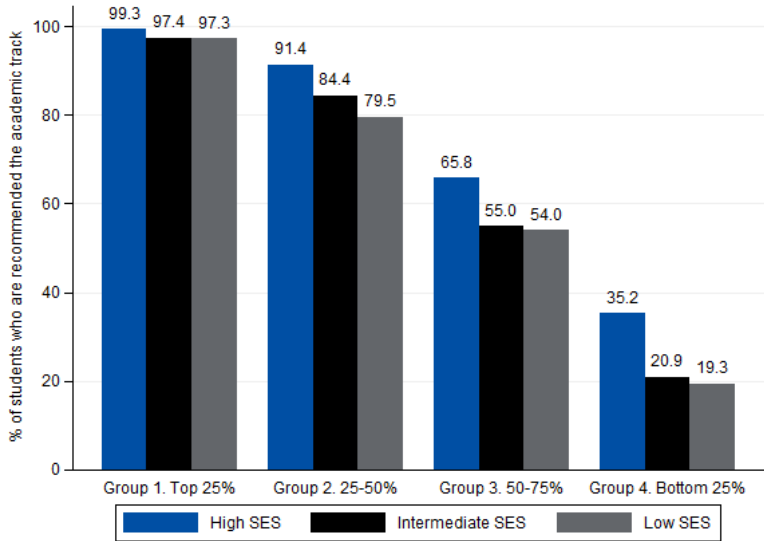


Figure 1.6: Track recommendations by performance in the final exam and SES

Figure 1.7 plots a similar graph for boys and girls. High-achieving boys and girls are almost always enrolled in the academic track. Though, at lower level of academic performance, the academic track recommendation rate is much higher for girls than for boys: in groups 2, 3 and 4, the share of girls who are recommended the academic track is respectively 9 pp., 11 pp. and 12 pp. higher than the one of boys.

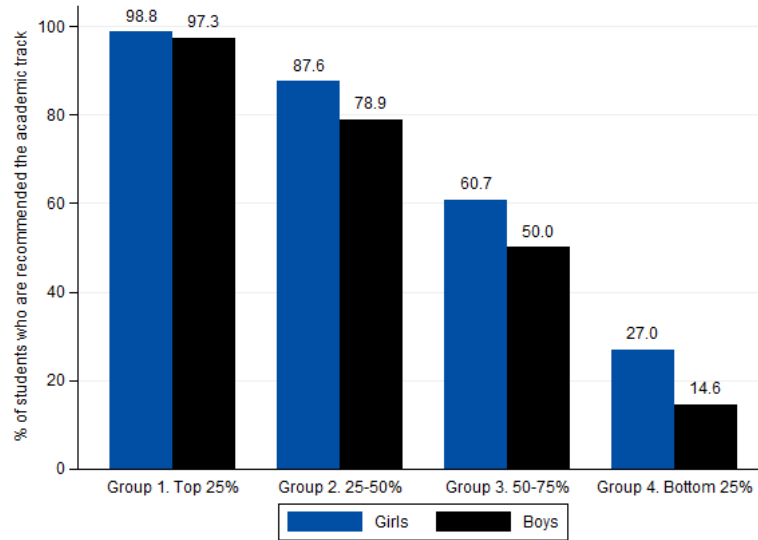


Figure 1.7: Track recommendations by performance in the final exam and Gender

### 1.5.2 Students' track choice independently of grading

In January, pupils indicate whether they intend to choose the academic track or a vocational track. We start by investigating whether this initial track choice depend on pupils' individual characteristics, independently of grading differentials.

**Model** Academic performance is the key factor that intervenes in the track choice because pupils choose a track that is in line with their potential. However, students do not observe their *true* academic ability. The best information they have about their ability is their score in in-class tests, which corresponds to their *perceived* ability  $\widetilde{\theta}_{ij}$ .

Other factors, that are not directly related to pupils' academic ability, might also affect their track choice. Duru-Bellat (1990) argues that schools transmit stereotypes on genders (i.e. some occupations are reserved to men). Boys and girls internalize these stereotypes and make track choices accordingly.

The SES may also play an important role on the initial track choice. As argued by Guyon & Huillery (2014), low-SES pupils might face more credit and liquidity constraints, which discourages them to choose the academic track because it involves a longer education period. They might also rationally anticipate that they have lower chance of succeeding in the academic track because their social and cultural environment is an handicap. They may be less informed about

the existing tracks, which could restrict their choice set. Their academic self-esteem may be lower if they do not have role models in their environment and if they internalize prejudices (stereotype susceptibility). Finally, peers may play an important role on the track choice: as low-SES pupils tend to be enrolled in disadvantaged schools, they may choose more often a vocational track if this is the social norm.

Pupils' track choice may also depend on their nationality, as students with immigrant backgrounds may have different aspirations (Caille (2007)).

We also need to include class fixed-effects in our model to account for neighborhood and peer effects. In order to simplify our results' interpretation, we define the track choice as a dummy variable that is equal to 1 when the student chooses the academic track and to 0 otherwise.<sup>14</sup> Our model is summarized in equation 1.2.

$$\begin{aligned} \textit{Academic track in initial choice}_{ij}^* &= \beta_0 + \beta_1 \widetilde{\theta}_{ij} + \beta_{girl} \textit{Girl}_i + \beta_{intSES} \textit{Intermediate SES}_i \\ &+ \beta_{lowSES} \textit{Low SES}_i + \beta_{foreign} \textit{Foreign}_i + \beta_{FE} \textit{FE}_j + u_{ij} \end{aligned} \quad (1.2)$$

where  $\textit{Academic track in initial choice}_{ij}^*$  is the latent variable associated to the propensity of student  $i$  in class  $j$  to choose the academic track and  $u_{ij}$  is random shock that follows a standard normal distribution.

**Estimation 1** Our aim is to estimate the  $\beta$  parameters from equation 1.2. We cannot estimate these parameters directly because the *perceived* academic ability  $\widetilde{\theta}_{ij}$  may be correlated with the error term  $u_{ij}$ . If a teacher likes a student, he might both grade him more favorably and support him more. This encouragement could boost the student's self-confidence and increase his likelihood of choosing the high-ability track. Such a mechanism would induce a correlation between  $\widetilde{\theta}_{ij}$  and  $u_{ij}$ .

Hence, we instrument  $\widetilde{\theta}_{ij}$  (the in-class score) by  $\theta_i$  (the final exam score) and we use our analysis of grading from equation 1.1 as the first stage of our IV estimation.

As the final exam score has a strong effect on the in-class score after partialling out the effect of all other regressors, the relevance condition of our instrument is satisfied (see table 1.3).

The exclusion condition of our instrument ( $cov(\theta_i, u_{ij}) = 0$ ) is also satisfied under our identifying

---

<sup>14</sup>Grouping together vocational tracks and grade repetition is not an issue in our setting for two reasons. First, grade repetition only represents 0.5% of the track choices so it should not alter the results. Second, at equivalent test scores, high-SES students and girls choose more often grade repetition. Hence our estimates of the effect of gender and SES on the probability of choosing the academic track should be a lower bound.

assumption. As  $\theta_i$  corresponds to pupils' *true* academic ability, it is orthogonal to teachers' biases and we can rule out the endogenous mechanism presented before.

We use a probit model to account for the fact that the effect of students' characteristics on the probability of choosing the academic track is larger for average achievers than for very low and very high achievers. We cluster standard errors at the class level to take into account potential correlation between unobservables of students in the same classroom.<sup>15</sup>

**Results** Average marginal effects of the IV probit estimation and their 95% confidence intervals<sup>16</sup> are reported in column (1) of table 1.5.

Table 1.5: Tracking (1/4) - average marginal effects from the probit estimation

<i>Dep. var:</i>	Initial choice	
In class score (in SD) <i>instrumented</i>	0.193*** [0.185,0.201]	
In class score (in SD) <i>non-instrumented</i>		0.246*** [0.239,0.252]
Girl	0.034*** [0.021,0.046]	0.025*** [0.013,0.037]
Intermediate SES	-0.085*** [-0.105,-0.066]	-0.088*** [-0.107,-0.069]
Low SES	-0.102*** [-0.120,-0.084]	-0.106*** [-0.124,-0.089]
Foreign	0.065** [0.025,0.104]	0.057** [0.016,0.097]
Class fixed-effects	Yes	Yes
Observations	16864	16864

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* corresponds to  $\beta_{girl}$  in the model (in a loose sense).

<sup>15</sup>We use the Stata command *ivprobit*.

<sup>16</sup>In our setting, reporting the average marginal effects makes more sense than reporting the marginal effect at the means because the effects of students' characteristics will be much greater at the average academic performance than at very high or very low level of academic performance. Hence, marginal effects at the means would not be representative of the overall effect.

The in-class score (instrumented) is significant at the 0.1% level and the magnitude of the effect is strong: when the in-class score increases by 1 SD, the probability of choosing the academic track increases by 19.3 percentage points (pp). This result shows that pupils make track choice that are (at least to some extent) coherent with their academic performance.

Students' characteristics included in the estimation are all statistically significant at the 1% level. Controlling for academic performance in class, an intermediate SES pupil and a low-SES pupil are respectively 8.5 and 10.2 pp. less likely to choose the academic track than a high-SES pupil. This finding is consistent with the existing literature on tracking in France, which provides evidence of lower aspirations for low-SES pupils when using teachers' grades (Davaillon & Nauze-Fichet (2004)). Gender and nationality also affects the initial track choice: girls apply more often than boys to the academic track (+ 3.4 pp), as already pointed out in the literature on tracking (see for example Vrignaud (2016)). Pupils with foreign nationality choose more often the academic track than French pupils with equivalent in-class score (+6.5 pp).

In column (2) of table 1.5, we report the average marginal effects obtained when we use the in-class score as a covariate without instrumenting it. The coefficients associated to gender, SES and nationality are not statistically different from those in the IV estimation. This suggests that the correlation between  $\widetilde{\theta}_{ij}$  and  $u_{ij}$  is very weak.

As a robustness check, we use a linear model instead of a probit (see table 1.12 in the Appendix). Gender, SES and nationality remain significant at the 1% level and the coefficients' signs do not change. Overall, our results show that these characteristics strongly affect pupils' track choice, independently of teachers' grading.

### 1.5.3 Impact of grading differentials on the track choice

Integrating our analysis on grading into our model of initial track choice allows us to understand how grading differentials indirectly impact pupils' track choice.

**Model** By combining equations 1.1 and 1.2, we get:

$$\begin{aligned}
 \textit{Academic track in initial choice}_{ij}^* &= \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1\theta_i + (\beta_{girl} + \beta_1\alpha_{girl})\textit{Girl}_i \\
 &+ (\beta_{intSES} + \beta_1\alpha_{intSES})\textit{Intermediate SES}_i + (\beta_{lowSES} + \beta_1\alpha_{lowSES})\textit{Low SES}_i \\
 &+ (\beta_{foreign} + \beta_1\alpha_{foreign})\textit{Foreign}_i + (\beta_{FE} + \beta_1\alpha_{FE})\textit{FE}_j + (u_{ij} + \beta_1\epsilon_{ij}) \quad (1.3)
 \end{aligned}$$

When considering pupils' initial track choice along with the final exam score, the coefficients associated to pupils' characteristics can be interpreted as the sum of the direct effect of the gap in initial track choice ( $\beta_X$ ), and the indirect effect of grading differentials on the initial track choice ( $\beta_1\alpha_X$ ). Thus, by comparing our results with the instrumented in-class score to our results with the final exam score, we can identify the indirect effect of grading differentials on the track choice.

**Estimation 2** We use a probit model to estimate parameters from equation 1.3. Our dependent variable is the pupil's initial track choice. Explanatory variables are the final exam score, dummy variables for pupils' characteristics and class fixed-effects. Standard errors are clustered at the classroom level.

**Results** Average marginal effects with the instrumented in-class score<sup>17</sup> and the final exam score are reported in table 1.6. We can interpret the difference between coefficients in column (2) and column (1) as the indirect effect of grading differentials on pupils' initial track choice.<sup>18</sup>

---

<sup>17</sup>This estimation corresponds to our previous analysis in column (1) of table 1.5

<sup>18</sup>Though, it is important to note that we report average marginal effects in this table. Hence, the coefficients in columns (1) and (2) do not exactly correspond to  $\beta_X$  and  $\beta_X + \beta_1\alpha_X$ . For the ease of interpretation, we do as if average marginal effects were equivalent to the model parameters in all our analysis on tracking. As all our conclusions remain valid when we use linear models - where the estimated coefficients exactly correspond to the models' parameters - this shortcut does not appear to be problematic (see table 1.14 in the Appendix for results with linear estimations).

Table 1.6: Tracking (2/4) - average marginal effects from probit estimations

<i>Dep. var:</i>		Initial choice
In class score (in SD) <i>instrumented</i>	0.193*** [0.185,0.201]	
Final exam score (in SD)		0.247*** [0.239,0.255]
Girl	0.034*** [0.021,0.046]	0.067*** [0.054,0.080]
Intermediate SES	-0.085*** [-0.105,-0.066]	-0.103*** [-0.122,-0.083]
Low SES	-0.102*** [-0.120,-0.084]	-0.122*** [-0.140,-0.103]
Foreign	0.065** [0.025,0.104]	0.083*** [0.043,0.122]
Class fixed-effects	Yes	Yes
Observations	16864	16864

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* in column (1) corresponds to  $\beta_{girl}$  in the model (in a loose sense).

Reading: the coefficient associated to *Girl* in column (2) corresponds to  $\beta_{girl} + \beta_1\alpha_{girl}$  in the model (in a loose sense).

The average marginal effect of gender on the initial track choice doubles when we consider the final exam score instead of the in-class score (from 3.4 pp to 6.7 pp). This means that about half of the gender gap in aspirations can be explained by the gender grading differential.

Marginal effects associated to SES and nationality increase when we consider the final exam score. However, confidence intervals in the two estimations overlap, which suggests that grading effects have a lower impact on the SES and nationality aspirations gap than on the gender aspirations gap. This outcome is coherent with the fact that we observe larger grading differentials between boys and girls than between low-SES and high-SES pupils or between French and foreign pupils (see table 1.3).

#### 1.5.4 Teachers' track recommendation independently of grading

In March, the teacher conference responds to the pupil's track choice (*Recommandations*), by either encouraging it or suggesting another track. Observing both pupils' initial track choice



and the teachers' track recommendation allows us to identify the direct influence of teachers on tracking.

**Model** Teachers take into account pupils' initial choice to make their track recommendations. Though, they do not systematically confirm the initial choice and they may play a corrective role. We define the latent variable associated to the propensity of student  $i$  in class  $j$  to be recommended the academic track as follows:

$$\text{Academic track recommended}_{i,j}^* = \text{Academic track in initial choice}_{i,j}^* + \text{Teachers' adjustment}_{i,j}$$

where according to equation 1.2:

$$\begin{aligned} \text{Academic track in initial choice}_{i,j}^* = & \beta_0 + \beta_1 \widetilde{\theta}_{i,j} + \beta_{girl} \text{Girl}_i + \beta_{intSES} \text{Intermediate SES}_i \\ & + \beta_{lowSES} \text{Low SES}_i + \beta_{foreign} \text{Foreign}_i + \beta_{FE} \text{FE}_j + u_{ij} \end{aligned}$$

and  $\text{Teachers' adjustment}_{i,j}$  is defined as follows:

$$\begin{aligned} \text{Teachers' adjustment}_{i,j} = & \gamma_0 + \gamma_1 \widetilde{\theta}_{i,j} + \gamma_{girl} \text{Girl}_i + \gamma_{intSES} \text{Intermediate SES}_i \\ & + \gamma_{lowSES} \text{Low SES}_i + \gamma_{foreign} \text{Foreign}_i + \gamma_{FE} \text{FE}_j + \eta_{ij} \end{aligned}$$

We assume that, controlling for pupil's initial track choice, the teachers' track recommendation depends on observed academic performance  $\widetilde{\theta}_{i,j}$ . Indeed, if a very high achieving student applies to a vocational track, teachers may recommend her the academic track instead. On the contrary, if a very low achieving student applies to the academic track, teachers may recommend her a vocational track instead.

Teachers' recommendation could also depend on pupils' characteristics. At equivalent initial choice and perceived academic performance, they might have prejudices towards certain categories of students and be less prone to recommend them the high-ability track. Another possible mechanism would be that teachers use their recommendations to "correct" the gender and SES gaps in initial track choice. In our setting, this would imply  $\gamma_{girl} < 0$  and  $\gamma_{lowSES} > 0$ .

Contextual specificity may also affect teachers' track recommendation. Like in our analysis on grading, we can conjecture that the degree of severity in track recommendations strongly depends on the class. Teachers in high-performing classes are likely to be more demanding than teachers in low-performing classes. Finally, the track recommendation depends on a random shock  $\eta_{ij}$ .

By combining equations on the initial track choice and on teachers' adjustment, we get the following expression:

$$\begin{aligned}
\textit{Academic track recommended}_{ij}^* &= \beta_0 + \gamma_0 + (\beta_1 + \gamma_1)\widetilde{\theta}_{ij} + (\beta_{girl} + \gamma_{girl})\textit{Girl}_i \\
&+ (\beta_{intSES} + \gamma_{intSES})\textit{Intermediate SES}_i + (\beta_{lowSES} + \gamma_{lowSES})\textit{Low SES}_i \\
&+ (\beta_{foreign} + \gamma_{foreign})\textit{Foreign}_i + (\beta_{FE} + \gamma_{FE})\textit{FE}_j + u_{ij} + \eta_{ij} \quad (1.4)
\end{aligned}$$

Thus, when considering teachers' track recommendation along with the in-class score, the coefficients associated to pupils' characteristics can be interpreted as the sum of the gap in initial track choice ( $\beta_X$ ) and the teachers' recommendation adjustment ( $\gamma_X$ ).

**Estimation 3** Our aim is to recover the  $\beta + \gamma$  parameters from equation 1.4. As mentioned previously,  $cov(\widetilde{\theta}_{ij}, u_{ij}) \neq 0$ . Furthermore, if a teacher likes a student, he might be less severe towards him through both grading and the track recommendation, which implies that  $cov(\widetilde{\theta}_{ij}, \eta_{ij}) \neq 0$ .

As  $\widetilde{\theta}_{ij}$  may be endogenous, we instrument it by  $\theta_i$ , as we did in our analysis of pupils' initial track choice. The exclusion condition is satisfied by the fact that  $\theta_i$  is independent of teachers' biases. We use a probit model and we cluster standard errors at the class level.

**Results** According to equations 1.2 and 1.4, we can identify the direct influence of teachers on tracking (i.e. the  $\gamma$  parameters) by comparing our coefficients in estimations 1 and 3. Obtaining the same coefficients means that teachers tend to simply confirm students' initial track choice whereas a statistical difference implies that teachers either accentuate or reduce gaps in the initial track choice. We report average marginal effects of the two estimations in columns (1) and (3) of table 1.7.

Table 1.7: Tracking (3/4) - average marginal effects from probit estimations

<i>Dep. var:</i>	Initial choice		Recommendation	
In class score (in SD) <i>instrumented</i>	0.193*** [0.185,0.201]		0.297*** [0.288,0.305]	
In class score (in SD) <i>non-instrumented</i>		0.246*** [0.239,0.252]		0.315*** [0.307,0.324]
Girl	0.034*** [0.021,0.046]	0.025*** [0.013,0.037]	0.026*** [0.016,0.037]	0.023*** [0.013,0.033]
Intermediate SES	-0.085*** [-0.105,-0.066]	-0.088*** [-0.107,-0.069]	-0.044*** [-0.061,-0.027]	-0.045*** [-0.062,-0.028]
Low SES	-0.102*** [-0.120,-0.084]	-0.106*** [-0.124,-0.089]	-0.054*** [-0.071,-0.038]	-0.056*** [-0.072,-0.040]
Foreign	0.065** [0.025,0.104]	0.057** [0.016,0.097]	0.049** [0.017,0.082]	0.046** [0.014,0.079]
Class fixed-effects	Yes	Yes	Yes	Yes
Observations	16864	16864	16864	16864

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* in columns 1-2 corresponds to  $\beta_{girl}$  in the model (in a loose sense).

Reading: the coefficient associated to *Girl* in columns 3-4 corresponds to  $\beta_{girl} + \gamma_{girl}$  in the model (in a loose sense).

The average marginal effect associated to the in-class score is larger when we consider the track recommendation than when we consider the initial track choice. An increase of 1 SD of the in-class score makes a pupil 19.3 pp more likely to apply to the academic track while it makes teachers 29.7 pp more likely to recommend the academic track. This result shows that teachers adjust the track recommendation according to pupils' observed academic performance ( $\gamma_1 > 0$ ). Pupils with bad in-class grades who apply to the academic track are recommended a vocational track (and vice-versa).

The coefficients associated to gender and nationality slightly change in the two estimations but the difference is not significant. Hence, teachers do not "correct" the gender and nationality gaps in initial track choice ( $\gamma_{girl} = 0$  and  $\gamma_{foreign} = 0$ ).

The coefficients associated to SES decrease by about 50% in the estimation of the track recommendation. An intermediate-SES (respectively low-SES) pupil is 8.5 pp (resp. 10.2 pp) less

likely to apply to the academic track whereas she is only 4.4 pp (resp 5.4 pp) less likely to be recommended the academic track. This result shows that teachers slightly reduce the initial social gap in aspirations by encouraging more low-SES pupils to pursue the academic track ( $\gamma_{intSES} > 0$  and  $\gamma_{lowSES} > 0$ ).

In column (4) of table 1.7, we report the marginal effects obtained when we use the in-class score as a covariate without instrumenting it in our analysis of teachers' recommendations. The coefficients associated to pupils' characteristics are almost the same as in the IV estimation. This suggests that the correlation between  $\widetilde{\theta}_{ij}$  and  $(u_{ij} + \eta_{ij})$  is very weak.

As a robustness check, we use a linear model instead of a probit model. All our conclusions remain valid (see table 1.13 in the Appendix).

### 1.5.5 Direct measures of teachers' impact on tracking

Integrating our analysis on grading into our model of teachers' track recommendation allows us to measure directly two effects: (i) the indirect effect of grading differentials on tracking through both pupil's initial track choice and teachers' track recommendation and (ii) the overall impact of teachers on tracking through both grading differentials and their track recommendations.

**Model** By combining equations 1.1 and 1.4, we get:

$$\begin{aligned}
\textit{Academic track recommended}_{ij}^* &= \beta_0 + \gamma_0 + (\beta_1 + \gamma_1)\alpha_0 + (\beta_1 + \gamma_1)\alpha_1\theta_i + [\beta_{girl} + \gamma_{girl} + (\beta_1 + \gamma_1)\alpha_{girl}]\textit{Girl}_i \\
&+ [\beta_{intSES} + \gamma_{intSES} + (\beta_1 + \gamma_1)\alpha_{intSES}]\textit{Intermediate SES}_i + [\beta_{lowSES} + \gamma_{lowSES} + (\beta_1 + \gamma_1)\alpha_{lowSES}]\textit{Low SES}_i \\
&+ [\beta_{foreign} + \gamma_{foreign} + (\beta_1 + \gamma_1)\alpha_{foreign}]\textit{Foreign}_i + [\beta_{FE} + \gamma_{FE} + (\beta_1 + \gamma_1)\alpha_{FE}]\textit{FE}_j + u_{ij} + \eta_{ij} + (\beta_1 + \gamma_1)\epsilon_{ij}
\end{aligned} \tag{1.5}$$

Thus, when considering the track recommendation along with the final exam score, the coefficients associated to pupils' characteristics can be interpreted as the sum of the gap in initial track choice ( $\beta_X$ ), the teachers' recommendation adjustment ( $\gamma_X$ ) and the indirect effect of grading differentials on tracking ( $(\beta_1 + \gamma_1)\alpha_X$ ).

**Estimation 4** We estimate parameters of equation 1.5 with a probit model. The dependent variable is the recommended track. Explanatory variables are the final exam score, dummy variables for pupil's characteristics and class fixed-effects. Standard errors are clustered at the

class level.<sup>19</sup>

**Results** According to equations 1.4 and 1.5, we can directly measure the indirect effect of grading differentials on tracking (i.e. the  $(\beta_1 + \gamma_1)\alpha_X$  parameters) by comparing our coefficients in estimations 3 and 4.

Furthermore, according to equations 1.2 and 1.5, we can directly measure the overall impact of teachers on tracking (i.e the  $\gamma_X + (\beta_1 + \gamma_1)\alpha_X$  parameters) by comparing our coefficients in estimations 1 and 4.

Average marginal effects from estimation 4 are reported in table 1.8 along with the three previous estimations.

---

<sup>19</sup>Note that we already estimated independently the different parameters of equation 1.5 in our previous analyses. Estimation 4 only allows us to measure directly the sum of the different effects, in a much more direct way than if we summed “manually” the different effects.

Table 1.8: Tracking (4/4) - average marginal effects from probit estimations

<i>Dep. var:</i>	Initial choice		Recommendation	
In class score (in SD) <i>instrumented</i>	0.193*** [0.185,0.201]		0.297*** [0.288,0.305]	
Final exam score (in SD)		0.247*** [0.239,0.255]		0.308*** [0.299,0.316]
Girl	0.034*** [0.021,0.046]	0.067*** [0.054,0.080]	0.026*** [0.016,0.037]	0.079*** [0.067,0.091]
Intermediate SES	-0.085*** [-0.105,-0.066]	-0.103*** [-0.122,-0.083]	-0.044*** [-0.061,-0.027]	-0.077*** [-0.096,-0.058]
Low SES	-0.102*** [-0.120,-0.084]	-0.122*** [-0.140,-0.103]	-0.054*** [-0.071,-0.038]	-0.088*** [-0.106,-0.070]
Foreign	0.065** [0.025,0.104]	0.083*** [0.043,0.122]	0.049** [0.017,0.082]	0.078*** [0.040,0.115]
Class fixed-effects	Yes	Yes	Yes	Yes
Observations	16864	16864	16864	16864

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* in column (1) corresponds to  $\beta_{girl}$  in the model (in a loose sense).

Reading: the coefficient associated to *Girl* in column (2) corresponds to  $\beta_{girl} + \beta_1\alpha_{girl}$  in the model (in a loose sense).

Reading: the coefficient associated to *Girl* in column (3) corresponds to  $\beta_{girl} + \gamma_{girl}$  in the model (in a loose sense).

Reading: the coefficient associated to *Girl* in column (4) corresponds to  $\beta_{girl} + \gamma_{girl} + (\beta_1 + \gamma_1)\alpha_{girl}$  in the model (in a loose sense).

Comparing columns (3) and (4) allows us to measure directly the effect of grading differentials on tracking. The average marginal effect associated to girl increases from 2.6 pp to 7.9 pp and the average marginal effect associated to low-SES goes from -5.4 pp to -8.8 pp. These results show that the gender and socio-economic tracking gaps are partly driven by teachers' grading differentials. Therefore, as postulated in the beginning of this paper, it is essential to investigate teachers' grading in any analysis on tracking.

Comparing columns (1) and (4) allows us to measure directly the overall impact of teachers on tracking through both their grading and their track recommendations. Teachers widens the initial gender gap in aspirations (the effect goes from 3.4 pp in estimation 1 to 7.9 pp in estimation 4). This is coherent with our previous findings: girls get higher in-class grades than

boys which indirectly affect tracking, and teachers do not use their recommendation to “correct downwards” the gender gap in the initial track choice. The impact of teachers on tracking with regards to pupils’ SES is more contrasted. Average marginal effects associated to SES are not statistically different in estimations 1 and 4. It means that globally, teachers do not widen nor reduce the socio-economic gap in aspirations. On one hand, their grading disadvantages low SES pupils, which indirectly widens the tracking gap. On the other hand, they use their track recommendations to slightly reduce the socio-economic gap in the initial track choice. These two antagonist effects seem to cancel each other.

When we use a linear model instead of a probit model, we obtain very similar results and all our conclusions are confirmed (see table 1.14 in the Appendix).

## 1.6 Discussion

In order to identify grading differentials, we assumed that the in-class-score reflects the *perceived* academic ability and that the final exam score reflects the *true* academic ability. In this section, we question further this identifying assumption and discuss its limits.

### 1.6.1 Interpreting the in-class score as the *perceived* academic ability

Students and teachers only observe in-class grades over the academic year, so it makes sense to consider that they reflect student’s *observable* or *perceived* academic ability. Nevertheless, data limitations regarding the in-class score and the assumption of common *perceived* academic ability by the student and her teachers need to be further discussed.

**Data limitations on the in-class score** A limit of our analysis is that we only have data on average in-class grades obtained over the entire academic year. As the initial track choice takes place in the end of the first quarter, ideally we should only consider the in-class score in the first quarter in our analysis of students’ aspirations. Similarly, as the track recommendation takes place in the end of the second quarter, ideally we should not take into account in-class grades from the third quarter in our analysis of teachers’ recommendations.

However, pupils’ academic performance is strongly correlated from one quarter to the next, so considering the in-class score over the entire year instead of the in-class score on more accurate periods should not strongly affect our results.

**Common *perceived* academic ability** We assume that the *perceived* academic performance of student  $i$  in class  $j$  is common to the student and her teachers and corresponds to the in-class score. Though, we could imagine that a student’s assessment of her own individual academic performance differs from the one of her teachers (and therefore does not correspond to her in-class score). For example, a student might consider that her in-class score under-estimates her performance. If such deviations between teachers’ and students’ perceptions were correlated to students’ individual characteristics, they could affect our results on tracking.

We cannot test empirically for the existence of such deviations, but from a theoretical viewpoint they seem very unlikely given the importance of grades in the French education system. Indeed, pupils and their parents tend to be obsessed by grades obtained at school (Merle (2014)). This obsession suggests that they think that grades reflect academic potential. Otherwise, they would not give them so much importance.

### 1.6.2 Interpreting the final exam score as the *true* academic ability

We now investigate two mechanisms according to which the final exam score would not necessarily reflect students’ *true* academic ability.

**The final exam is “one-shot”** The final exam score is based on academic performance in a one shot exam, so one concern might be that performance in this test varies at the individual level due to random variations around average ability (for example if the student is sick during the exam day). Considering these random variations are orthogonal to pupils’ individual characteristics, our results are not biased. Only the variance of our estimates could be affected. As most of our results are significant at the 0.1% level, this should not affect any of our conclusions.

**The final exam is a high-stake situation** The final exam is more stressful than in-class tests because it is weighted much more than an in class test in the overall score. Moreover, the final test is the first official exam that students take in their life, so pressure might be pretty high for them. If students respond differently to pressure depending on their gender or social background, their final exam score may not only reflect their *true* academic ability, but also a “choking under pressure” component that is correlated to gender and SES. Azmat, Calsamiglia & Iriberry (2016) test for such an effect by gender and show that female students outperform male students relatively more when the stakes are low.



In order to ensure that our results on grading differentials are not driven by choking under pressure, we focus on a subsample of students who do not face pressure in the final exam. As in-class tests account for 60% of the total number of points required to get the DNB degree, some very high-performing pupils do not need any point in the final exam to pass.<sup>20</sup> Hence, we can assume that for these top students, different responses to pressure cannot explain grading differentials by gender and SES.

We perform the same estimations as in section 1.4 on this subsample of pupils. Results are reported in table 1.9. Grading differentials associated to gender and low-SES remain significant at the 5% level. This is evidence that choking under pressure does not (at least not fully) explain grading differentials.

Table 1.9: Grading differentials (subsample of top students with no pressure)

<i>Dep. var:</i>	In-class score (in SD)	
Final exam score (in SD)	0.423*** (0.023)	
Final exam score (in pctiles)		Yes
Girl	0.045* (0.020)	0.043* (0.021)
Intermediate SES	-0.023 (0.020)	-0.030 (0.021)
Low SES	-0.047* (0.021)	-0.049* (0.022)
Foreign	-0.186** (0.070)	-0.212*** (0.058)
Constant	1.355*** (0.044)	2.430*** (0.036)
Class fixed-effects	Yes	Yes
Observations	1523	1523
$R^2$	0.70	0.72

Standard errors are clustered at the class level and are reported in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

<sup>20</sup>In our setting, 1523 students have enough points in-class to get the DNB degree without getting any point in the final exam. They represent 9.2% of the overall sample.

## 1.7 Conclusion

This paper provides compelling evidence that grading and tracking are two interconnected questions in education. As grades are supposed to reflect pupils' academic potential and their chance of succeeding in the different tracks, teachers' grading indirectly affect tracking.

We investigate teachers' grading practices by comparing students' performance in a blind standardized final exam that reflects pretty well their *true* academic potential to their performance in non-blind class tests that are left at teachers' discretion. At equivalent score in the final exam, boys and low-SES students get lower grades in class.

This analysis on grading allows us to isolate direct tracking effects from indirect grading effects. We show that boys and low-SES pupils apply much less often to the academic track than their equally achieving classmates and that teachers' grading differentials amplify these gender and socio-economic gaps in aspirations.

# Appendices

## Appendix A: defining students' SES

In our dataset, we have detailed information on parents' SES according to the French education classification, which builds on the National Statistics Institute (*INSEE*) classification. Each category of occupation is assigned a 2-digit code. We follow the same methodology as Guyon & Huillery (2014) to classify occupations in three categories:

- *High-SES occupations* correspond to codes 23, 31, 33, 34, 35, 37 and 38.
- *Intermediate-SES occupations* are referred to codes 42, 43, 44, 45, 46, 47, 48 and 73.
- *Low-SES occupations* correspond to codes 10, 21, 22, 52, 53, 54, 55, 56, 61, 66, 69, 71, 72, 76, 81, and 82.

The student SES is then defined according to her parents' occupations:

- If at least one of her parent has a high-SES occupation, the student is assigned to the category "high-SES".
- If her two parents have low-SES occupations, the student is assigned to the category "low-SES".
- In all other cases (except when information is missing for both parents, in which case we exclude the student from the sample), the student is assigned to the category "intermediate SES".

		Parent 1			
		High-SES	Intermediate SES	Low-SES	Missing data
Parent 2	High-SES	H	H	H	H
	Intermediate SES	H	I	I	I
	Low-SES	H	I	L	L
	Missing data	H	I	L	

H = high-SES student, I = intermediate SES student, L = low-SES student

Table 1.10: Defining students' SES

## Appendix B: Additional tables

Table 1.11: Grading differentials by subject

<i>Dep. var:</i> In-class score (in SD)	(1)	(2)	(3)	(4)	(5)	(6)
	Mathematics		History/Geography		French	
Final exam score (in SD)	0.846*** (0.006)		0.784*** (0.006)		0.753*** (0.006)	
Final exam score (in pctiles)		Yes		Yes		Yes
Girl	0.173*** (0.010)	0.170*** (0.010)	0.137*** (0.011)	0.135*** (0.011)	0.172*** (0.011)	0.172*** (0.011)
Intermediate SES	-0.102*** (0.013)	-0.104*** (0.013)	-0.145*** (0.014)	-0.146*** (0.014)	-0.196*** (0.015)	-0.197*** (0.015)
Low SES	-0.126*** (0.012)	-0.128*** (0.012)	-0.169*** (0.013)	-0.169*** (0.013)	-0.246*** (0.014)	-0.246*** (0.014)
Foreign	0.032 (0.035)	0.035 (0.035)	0.020 (0.037)	0.017 (0.037)	0.023 (0.042)	0.027 (0.042)
Constant	-0.107*** (0.011)	-1.721*** (0.037)	0.213*** (0.012)	-1.513*** (0.046)	-0.002 (0.012)	-1.714*** (0.048)
Class fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16864	16864	16864	16864	16864	16864
$R^2$	0.74	0.74	0.72	0.72	0.68	0.68

Standard errors are clustered at the class level and are reported in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 1.12: Tracking - pupils' initial track choice - linear estimations

<i>Dep. var:</i>	Initial choice	
In class score (in SD) <i>instrumented</i>	0.261*** [0.254,0.269]	
In class score (SD) <i>non-instrumented</i>		0.252*** [0.245,0.259]
Girl	0.021** [0.008,0.034]	0.023*** [0.010,0.036]
Intermediate SES	-0.035*** [-0.051,-0.019]	-0.040*** [-0.057,-0.023]
Low SES	-0.055*** [-0.070,-0.040]	-0.061*** [-0.077,-0.046]
Foreign	0.071** [0.026,0.116]	0.069** [0.024,0.115]
Constant	1.093*** [1.080,1.106]	1.095*** [1.081,1.108]
Class fixed-effects	Yes	Yes
Observations	16864	16864
$R^2$	0.39	0.40

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* corresponds to  $\beta_{girl}$  in the model.

Table 1.13: Tracking - pupils' initial track choice vs. teachers' track recommendation - linear estimations

<i>Dep. var:</i>	Initial choice		Recommendation	
In class score (in SD) <i>instrumented</i>	0.261*** [0.254,0.269]		0.323*** [0.316,0.330]	
In class score (in SD) <i>non-instrumented</i>		0.252*** [0.245,0.259]		0.331*** [0.324,0.337]
Girl	0.021** [0.008,0.034]	0.023*** [0.010,0.036]	0.019*** [0.008,0.030]	0.017** [0.006,0.029]
Intermediate SES	-0.035*** [-0.051,-0.019]	-0.040*** [-0.057,-0.023]	-0.012 [-0.028,0.003]	-0.008 [-0.024,0.008]
Low SES	-0.055*** [-0.070,-0.040]	-0.061*** [-0.077,-0.046]	-0.023** [-0.037,-0.008]	-0.017* [-0.032,-0.003]
Foreign	0.071** [0.026,0.116]	0.069** [0.024,0.115]	0.057** [0.016,0.098]	0.058** [0.016,0.100]
Constant	1.093*** [1.080,1.106]	1.095*** [1.081,1.108]	0.721*** [0.709,0.734]	0.720*** [0.707,0.733]
Class fixed-effects	Yes	Yes	Yes	Yes
Observations	16864	16864	16864	16864
$R^2$	0.39	0.40	0.53	0.53

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* in columns 1-2 corresponds to  $\beta_{girl}$  in the model.

Reading: the coefficient associated to *Girl* in columns 3-4 corresponds to  $\beta_{girl} + \gamma_{girl}$  in the model.

Table 1.14: Tracking - summary table - linear estimations

<i>Dep. var:</i>	Initial choice		Recommendation	
In class score (in SD) <i>instrumented</i>	0.261*** [0.254,0.269]		0.323*** [0.316,0.330]	
Final exam score (in SD)		0.237*** [0.230,0.245]		0.293*** [0.286,0.300]
Girl	0.021** [0.008,0.034]	0.064*** [0.050,0.077]	0.019*** [0.008,0.030]	0.072*** [0.059,0.085]
Intermediate SES	-0.035*** [-0.051,-0.019]	-0.057*** [-0.074,-0.040]	-0.012 [-0.028,0.003]	-0.040*** [-0.057,-0.023]
Low SES	-0.055*** [-0.070,-0.040]	-0.080*** [-0.096,-0.064]	-0.023** [-0.037,-0.008]	-0.054*** [-0.070,-0.038]
Foreign	0.071** [0.026,0.116]	0.086*** [0.039,0.133]	0.057** [0.016,0.098]	0.076** [0.029,0.124]
Constant	1.093*** [1.080,1.106]	1.092*** [1.078,1.106]	0.721*** [0.709,0.734]	0.721*** [0.707,0.734]
Class fixed-effects	Yes	Yes	Yes	Yes
Observations	16864	16864	16864	16864
$R^2$	0.39	0.35	0.53	0.43

Standard errors are clustered at the class level

95% confidence intervals in brackets

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reading: the coefficient associated to *Girl* in column (1) corresponds to  $\beta_{girl}$  in the model.

Reading: the coefficient associated to *Girl* in column (2) corresponds to  $\beta_{girl} + \beta_1\alpha_{girl}$  in the model.

Reading: the coefficient associated to *Girl* in column (3) corresponds to  $\beta_{girl} + \gamma_{girl}$  in the model.

Reading: the coefficient associated to *Girl* in column (4) corresponds to  $\beta_{girl} + \gamma_{girl} + (\beta_1 + \gamma_1)\alpha_{girl}$  in the model.

## Chapter 2

# Individual contribution in team contests

ANTOINE CHAPSAL & JEAN-BAPTISTE VILAIN



### **Abstract**

This paper empirically analyzes team effects in multiple pairwise battles, where players from two rival teams compete sequentially. Using squash team contests as a randomized natural experiment, we show that winning the first battle significantly increases the probability of winning the subsequent one. We derive testable predictions from a theoretical model in order to identify the effect driving this dependence on past outcomes. We provide compelling evidence of an *individual contribution effect*: players not only benefit from their team's win, but also value the fact of being individually – even partly – responsible for their team's collective success. Such an effect is of prime importance to understanding why individuals can make a significant effort when offered collective-based incentives.

**JEL classification:** C72, D79, L83, M54.

**Keywords** Teams; Multiple Pairwise Battles; Individual Contribution.

## 2.1 Introduction

Many economic achievements are produced by groups, such as teams or partnerships, in which each individual outcome mainly depends on other teammates' effort decisions. This is why there is an extensive body of economic literature devoted to understanding individuals' behavior in teams and exploring the design of team-based incentives. The conventional economic wisdom is that team-based incentives induce individuals to exert less effort than individual-based contracts, simply because rational and self-interested individuals free ride and do not internalize their teammates' utility when making effort decisions.<sup>1</sup> However, recent literature has found that individuals could make a significant effort for their team, not only when they react to peer pressure, but also in order to avoid feeling guilty, i.e., living up to the expectations of others (Kandel & Lazear (1992), Charness & Dufwenberg (2006), Chen & Lim (2013)). In this paper, we provide compelling evidence that individuals value being at least partly responsible for their team's success. Teammates make a significant effort in teams because they want to take part in the group's success. We refer to this important team effect as *individual contribution*: when involved within a team, individuals positively take into account the role they can play to achieve collective success.

Assessing individual contribution to team output is extremely challenging, especially when teammates' efforts interact in a subtle manner (Alchian & Demsetz (1972)).<sup>2</sup> We focus on a special kind of team setting, "multiple pairwise battles," where individual production is fully observable, thereby allowing us to test whether individuals value contributing to their team's victory. Multiple pairwise battles<sup>3</sup> refer to collective contests where players from two rival teams compete in individual battles and the winning team is the one that wins the majority of battles. Such contests correspond to many economic and social phenomena. For instance, in the House of Representatives and Senate elections in the United States, rival political parties compete for legislative seats in each electoral district, and a party can form a government or set political agenda in the legislature if it acquires majority status. An R&D race for a new product can also be seen as multiple pairwise battles: it implies competition on a series of component technologies

---

<sup>1</sup>See Prendergast (1999) and Sheremeta (2017) for surveys on this topic.

<sup>2</sup>The econometric estimation of individual productivity in team settings is investigated in the third chapter of this dissertation.

<sup>3</sup>The expression "multiple pairwise battles" is used by Fu, Lu & Pan (2015). The alternative expression "multi-battle team contest" is also used by some authors. In this paper, we refer to each component of a contest as a battle, a match or a confrontation.

and the winning firm is the one that gets ahead of its competitors in the majority of technologies. Some sports events also correspond to this kind of contest. A famous example is the Davis Cup tennis tournament, where the players from two national teams compete sequentially in a best-of-five contest.

Fu, Lu & Pan (2015) present a benchmark theoretical analysis of multiple pairwise battles. They show, under standard assumptions, that the outcome of a battle is independent from the outcome of previous and subsequent confrontations. Such a result, which they refer to as “neutrality,” implies that confrontations can be considered independent. There is not any “dynamic linkage” between subsequent battles and the order of play does not affect the final result. Neutrality is derived from the fact that players do not internalize the cost of effort of upcoming battles, for the simple reason that it is borne by their teammates.

From a theoretical viewpoint, one can draw a contrast with the “discouragement” effect, which arises in *individual* multi-battle contests in which the same players square off against one another sequentially.<sup>4</sup> In individual contests (e.g., a two-set tennis match), winning the first confrontation (or the first set) positively affects the probability of winning the next one: the remaining effort required to win the contest is lower for the frontrunner than for the laggard. The former is therefore more likely to win than the latter. The discouragement effect, which has been studied extensively,<sup>5</sup> cannot occur in multiple pairwise battles, as the remaining effort required to obtain the final payoff after a non-definitive battle is not to be borne by the current player.

As Fu, Lu & Pan (2015) stress, the neutrality result contrasts sharply with conventional wisdom, which holds that battles are not independent in a team contest. From a theoretical perspective, two kinds of effects would explain why winning the first battle should affect the outcome of the subsequent one in multiple pairwise battles: (i) effects that endogenously alter

---

<sup>4</sup>See Dechenaux, Kovenock & Sheremeta (2015) for a survey.

<sup>5</sup>There is an abundant literature on individual contests, which finds evidence of the dependence of outcomes in subsequent individual confrontations and confirms the discouragement effect. For instance, Klumpp & Polborn (2006) model U.S. presidential primaries as a best-of-N contest between two candidates and show that winning the early districts strongly affects the probability of winning later districts. Malueg & Yates (2010) find empirical evidence of strategic effects in individual tennis matches. Taking a sample of equally skilled players, they show that the winner of the first set exerts more effort in the second set than the loser. Mago, Sheremeta & Yates (2013) provide experimental evidence of a discouragement effect in a best-of-three Tullock contest. They also show that this effect is strategic, not psychological. Harris & Vickers (1987) show that in a two-firm R&D race model, an early lead yields easy wins in subsequent battles because of the discouragement effect on the lagging opponent. Konrad & Kovenock (2009) show in a theoretical framework that the introduction of intermediate prizes for component battles (i.e., payoff from winning a single battle even if the match is lost) reduces discouragement.

players' ability; and (ii) effects that endogenously affect players' incentives.

First, two effects, "choking under pressure" and "psychological momentum," may explain the absence of neutrality as they alter players' ability. Choking under pressure may occur if a player in the lagging team faces more pressure than a player in the leading team because the former has to win in order to keep his team in the contest. This pressure might have a detrimental effect on performance and might thus explain why winning the first battle would affect the probability of winning the next one. Psychological momentum, which has been mainly documented in individual settings, denotes the idea that winning a battle boosts players' confidence and helps them win the next one. In other words, initial success in a contest produces momentum that leads to future success.

Second, outcome dependence in multiple pairwise battles may be explained by asymmetric incentives among players, which may be caused by three potential phenomena. Altruism, i.e., the fact that players internalize their teammates' costs of effort, would generate a linkage between subsequent battles. If individuals were altruistic, bearing part of their teammates' costs, the player in the leading team would have more incentives to win than the player in the lagging team because he could avoid one of his teammate to make a costly effort in the upcoming battle. Another effect that may distort players' incentives and generate outcome dependence is "guilt aversion": a player may dread being (partly) responsible for his team's defeat. In this case, the player in the lagging team has more incentives to win than his opponent because he is more likely to be responsible for his team's defeat than his opponent if he loses his individual battle. Note that contrary to other effects, guilt aversion implies that being in the lagging team increases the probability of winning the battle. Finally, one may also consider another kind of effect, which we refer to as "individual contribution": players may value being partly responsible for collective success. In such a case, a player on the leading team has a higher probability of being (partly) responsible for collective success than his opponent. This higher probability increases his incentive to make a more costly effort, thereby increasing his probability of winning. It would follow, then, that winning the first battle endogenously creates asymmetry in incentives and may therefore lead to outcome dependence with subsequent battles.

These potentially strong effects are grounds to empirically test for neutrality, which consists in analyzing whether winning the first battle affects the probability of winning the second one. International team squash contests appear as an ideal laboratory to answer this question. They exactly correspond to best-of-three multiple pairwise battles: three players from two rival teams

compete sequentially, with each player playing only once and a team needs to win two individual battles in order to win the contest. Furthermore, the sequence of individual battles in a team contest is randomly drawn and cannot be manipulated.

We use this unique randomized empirical setting in order to test for the neutrality result on field data and identify the effects at stake. Our contribution is twofold. First, we find evidence of a dynamic linkage between subsequent battles. More precisely, we show that, *ceteris paribus*, winning the first battle significantly increases the probability of winning the second battle. This team effect contradicts neutrality. Second, we derive testable predictions from a theoretical model to further explain outcome dependence in this team setting and identify the effect at play. We provide compelling evidence that the predictions of the individual contribution effect are all validated in the data: (i) Winning battle 1 increases the probability of winning battle 2 (non-neutrality). This is coherent with the fact that the player in the leading team has more incentives to win battle 2 than his opponent because he is sure to contribute to the success of his team if he wins whereas the player in the lagging team will get the contribution reward if he wins only if his teammate also wins battle 3. Hence, there is an asymmetry of contribution incentives ; (ii) The probability of winning battle 2 increases with the probability of winning battle 3. The intuition behind this prediction is that the asymmetry of contribution incentives in battle 2 depends on the expected outcome of battle 3. For example, if the player in the lagging team in battle 2 knows that his teammate will lose battle 3, he knows that his team will lose the contest and that he will not get the contribution reward, whatever the outcome of his individual battle. In this case, the asymmetry of contribution incentives reaches its maximum ; (iii) Winning battle 1 and battle 2 increases the probability of winning trivial battle 3. This prediction is based on the idea that the individual contribution effect should also be at play in a trivial battle 3 - where, by definition, the winning team has already been determined. A player involved in a trivial battle 3 whose team has already won the contest should also value winning his match, so as to take part, albeit less directly than his teammates, in the success of his team. Hence he has more incentives to win than his opponent.

These three empirical findings support the fact that players value being (at least partly) responsible for collective success. On the contrary, theoretical predictions from the other potential effects at stake (choking under pressure, psychological momentum, altruism and guilt aversion) are not supported by the data.

Our results are robust to several specifications and alternative tests. To the best of our

knowledge, this is the first paper to provide field evidence that teammates individually value contributing to their team’s success. The implications of our results extend beyond contest designs. Indeed, the individual contribution effect is of prime importance to understanding why team-based incentives induce individuals to make a significant effort. It appears to be one of the factor explaining why forming teams is economically desirable.

The remainder of this paper is organized as follows. Section 2.2 presents the related literature. Section 2.3 provides empirical evidence against neutrality in multiple pairwise battles: winning the first battle significantly increases the probability of winning the second battle. Section 2.4 pinpoints the mechanism driving non-neutrality: outcome dependence is consistent with the fact that players value not only the final reward yielded by their team’s win, but also playing an active part in collective success. We also present robustness tests that confirm the existence of individual contribution in multiple pairwise battles and rule out alternative explanations. Section 2.5 concludes the paper with a discussion of the main implications of the individual contribution effect.

## 2.2 Related literature

A few papers directly test for neutrality. Fu, Ke & Tan (2015) conduct a simple best-of-three team contest experiment, in which players from two rival teams are pairwise matched and compete by counting the number of zeros in a series of 10-digit number strings composed of 0s and 1s. They find evidence that players from both teams remain equally motivated after observing the outcome of the first component contest, and therefore a team tournament is equally likely to end after two or three component contests. Huang (2016) uses team squash data and does not find evidence against neutrality. His findings are based on a limited number of matches. Huang & Murad (2017) develop an experiment to test for neutrality in a sequential best-of-three team contest. In their experiment, subjects have a fixed amount of time to catch balls that fall from the top of a computer screen by using mouse clicks. In a benchmark environment where there is no communication amongst teammates, who only share the same fate in terms of financial incentives, they find that second movers in lagging positions drop out of competition more often than second movers in leading positions. This result contradicts neutrality and can be supported by a psychological momentum. However, when teammates chat at the beginning of contests, the outcomes of the first two battles are independent. Huang & Murad (2017)’s interpretation

is that guilt aversion compensates the psychological effect observed in the benchmark situation.

Most of these studies rely on experimental data from controlled settings. Testing for neutrality on field data poses two additional empirical challenges.

First, failure to account for peer effects would generate a spurious correlation between the outcomes of subsequent battles. Being in a more stimulating environment might increase each teammate's probability of winning, thereby generating some spurious correlation between the probabilities of victory in the first two battles. The existence of peer effects continues to be debated in the literature. For instance, Mas & Moretti (2009) show, using high-frequency data from a field experiment, that the introduction of highly productive personnel into a team has a positive effect on worker productivity. On the contrary, Guryan, Kroft & Notowidigdo (2009) find no evidence of peer effects in a highly skilled professional labor market: neither the ability nor the current performance of playing partners affects the performance of professional golfers. In our setting, we can account for potential peer effects by including teams' rankings as control variables in our estimation. Teams' rankings actually encompass most of the environment effects that may be at play, including peer effects.

Second, belonging to the lagging team should not induce the player involved in the second battle to adopt a particular strategy that would affect the outcome of his match. The literature has focused on the fact that players could adopt riskier strategies when facing critical situations. Knoeber & Thurman (1994) compare tournament and linear payment schemes using data from a sample of U.S. broiler producers. They examine the impact of prizes on performance level and variability, concluding that less able producers adopt riskier strategies. On the contrary, Brown (2011), who shows that professional golfers underperform when they are paired with a superstar, concludes that this reduced performance is not attributable to the adoption of risky strategies. In our setting, we can test whether players adopt riskier strategies in critical situations by comparing the characteristics of battle 1 and battle 2 matches such as duration. We do not find statistical difference between the two periods, which suggest that players do not adopt risky strategies (see section 2.3.2).

## 2.3 Testing for neutrality in multiple pairwise battles

### 2.3.1 Theoretical framework

This section theoretically analyzes individual behaviors in multiple pairwise battles and presents the equilibrium probability of winning a component battle in a tractable form, which allows us to (i) show that the neutrality result from Fu, Lu & Pan (2015) leans on two important assumptions, and (ii) derive predictions for the empirical analysis (see section 2.4).

#### Setting

We consider a best-of-three team contest with complete information. A team  $X$  is opposed to a team  $Y$ . The contest presents the following features: (i) there are 3 risk-neutral players in each team. Each player only plays one battle.  $X_i$  (respectively  $Y_i$ ) is the player from team  $X$  (respectively  $Y$ ) that plays the  $i^{th}$  battle,  $i = \{1, 2, 3\}$ ; (ii) team  $X$  wins as soon as it wins two battles and loses as soon as it loses two battles; and (iii) the third battle is non-trivial only if team  $X$  and team  $Y$  have both won one of the two previous battles.

Let  $U_{X_i}$  and  $U_{Y_i}$  be the respective utilities of players  $X_i$  and  $Y_i$ .

Let  $p_i$  be the probability that  $X_i$  wins his battle against  $Y_i$ ,

$$p_i = \frac{x_i}{x_i + y_i},$$

where  $x_i$  is the level of effort of  $X_i$  and  $y_i$  is the level of effort of  $Y_i$ . This function is the simplest version of the Tullock contest success function,<sup>6</sup> also referred to as a lottery contest. Players do not have the same ability. This is reflected in a linear cost function, given by

$$C_{X_i} = \frac{x_i}{\theta_{X_i}},$$

where  $\theta_{X_i}$  is the innate ability of  $X_i$ . The cost of effort is thus a decreasing function of the innate ability of a player. The payoff associated with the collective win (denoted  $V$ ) is the same for every player. Players also get a battle reward  $v$  when they win their own battle (independently of their team's outcome).  $V$  and  $v$  are strictly positive.

#### Theoretical result

**Result 1. *Equilibrium probability of winning.*** *In a multiple pairwise battle, players choose their optimal level of effort such that the probability that player  $X_i$  wins battle  $i$  ( $i$  being any of*

---

<sup>6</sup>See Buchanan, Tollison & Tullock (1980).



the three individual battles in the team contest) is given by

$$p_i^* = \frac{\theta_{X_i} \Delta_{U_{X_i}}}{\theta_{X_i} \Delta_{U_{X_i}} + \theta_{Y_i} \Delta_{U_{Y_i}}},$$

where  $\Delta_{U_{X_i}} = (U_{X_i}|Win_{X_i}) - (U_{X_i}|Loss_{X_i})$ , respectively  $\Delta_{U_{Y_i}} = (U_{Y_i}|Loss_{X_i}) - (U_{Y_i}|Win_{X_i})$ , is the prize spread of player  $X_i$ , respectively  $Y_i$ .

*Proof.* Let  $U_{J_i}|Win_{K_i}$  (respectively  $U_{J_i}|Loss_{K_i}$ ) be the utility of player  $J_i$ ,  $J_i = \{X_i, Y_i\}$ , when  $K_i$  wins (respectively loses) battle  $i$ ,  $K_i = \{X_i, Y_i\}$ .

Both players choose their level of effort to maximize their expected utility:

$$\begin{aligned} \max_{x_i} & \left( \frac{x_i}{x_i + y_i} (U_{X_i}|Win_{X_i}) + \frac{y_i}{x_i + y_i} (U_{X_i}|Loss_{X_i}) - \frac{x_i}{\theta_{X_i}} \right), \\ \max_{y_i} & \left( \frac{y_i}{x_i + y_i} (U_{Y_i}|Loss_{X_i}) + \frac{x_i}{x_i + y_i} (U_{Y_i}|Win_{X_i}) - \frac{y_i}{\theta_{Y_i}} \right). \end{aligned}$$

Assuming that  $U$ s are independent of  $x_i$  and  $y_i$ , the first order conditions yield the following optimal levels of effort and equilibrium probability of winning  $p_i^*$ :

$$\begin{aligned} x_i^* &= \frac{(\theta_{X_i} \Delta_{U_{X_i}})^2 \theta_{Y_i} \Delta_{U_{Y_i}}}{(\theta_{X_i} \Delta_{U_{X_i}} + \theta_{Y_i} \Delta_{U_{Y_i}})^2}, \\ y_i^* &= \frac{\theta_{X_i} \Delta_{U_{X_i}} (\theta_{Y_i} \Delta_{U_{Y_i}})^2}{(\theta_{X_i} \Delta_{U_{X_i}} + \theta_{Y_i} \Delta_{U_{Y_i}})^2}. \end{aligned}$$

Finally,

$$p_i^* = \frac{\theta_{X_i} \Delta_{U_{X_i}}}{\theta_{X_i} \Delta_{U_{X_i}} + \theta_{Y_i} \Delta_{U_{Y_i}}}.$$

where  $\Delta_{U_{X_i}} = (U_{X_i}|Win_{X_i}) - (U_{X_i}|Loss_{X_i})$  – respectively,  $\Delta_{U_{Y_i}} = (U_{Y_i}|Loss_{X_i}) - (U_{Y_i}|Win_{X_i})$ , denotes player  $X_i$ 's – respectively, player  $Y_i$ 's prize spread.  $\square$

This result shows that the outcome of a battle depends on two parameters only, which are (i) players' relative prize spreads, and (ii) players' relative ability (or cost of effort).

Therefore, if players have the same prize spreads (i.e.,  $\Delta_{U_{X_i}} = \Delta_{U_{Y_i}}$ ) and a cost of effort that does not depend on the circumstances of the battle (i.e.,  $\theta_{X_i}$  and  $\theta_{Y_i}$  are not affected by the state of the contest), the equilibrium probability of winning only depends on players' relative ability:

$$p_{iNeutrality}^* = \frac{\theta_{X_i}}{\theta_{X_i} + \theta_{Y_i}}.$$

In this case, the team contest boils down to a series of independent lotteries, yielding Fu, Lu & Pan (2015)'s neutrality result according to which the outcome of a battle does not affect the outcome of the subsequent ones (i.e., leading or lagging behind has no effect).

Neutrality is thus based on two crucial assumptions, which are (i) common prize spreads, and (ii) the absence of phenomena altering players' effort cost or ability.

**(i) Common prize spreads** We observe neutrality (i.e.,  $p_i^* = \frac{\theta_{X_i}}{\theta_{X_i} + \theta_{Y_i}}$ ) if and only if players have common prize spreads (i.e.,  $\Delta_{U_{X_i}} = \Delta_{U_{Y_i}}$ ). This condition is satisfied in the case where players only value the collective win (payoff  $V$ ) and the battle reward (payoff  $v$ ). In a decisive battle 3,<sup>7</sup> both players have a prize spread of  $V + v$ , as they get both the collective and the battle rewards if they win and a payoff of 0 if they lose. In battle 2, both players also have the same prize spread: the player in the leading team gets  $V + v$  if he wins, and  $p_3^*V$  if he loses (as he can still get the collective reward  $V$  if his teammate wins battle 3, which occurs with a probability  $p_3^*$ ), so his prize spread is  $V + v - p_3^*V = v + (1 - p_3^*)V$ . The player in the lagging team gets  $v + (1 - p_3^*)V$  if he wins, as he is certain to receive the battle reward and he also gets the collective reward if his teammate wins battle 3, which occurs with a probability  $(1 - p_3^*)$ . If he loses, the contest ends and he gets a payoff 0, so his prize spread is also  $v + (1 - p_3^*)V$ . A similar logic applies to battle 1.

A necessary condition for players to have common prize spreads is that they do not take into account their teammates' costs of effort. A situation where players would, for any reason, act in an altruistic way and internalize part of their teammates' effort cost would generate asymmetric prize spreads in battle 2. This would explain why there could exist a dynamic linkage between subsequent battles.

Furthermore, players might not only value the team win and the battle reward but also being individually (partly) responsible for collective success. If such motivation exists, the prize spreads become asymmetric in battle 2: the player in the leading team has more incentive to win than his opponent because he is sure to contribute to the success of his team if he wins his battle, while his opponent will be "success-responsible" if and only if his teammate also wins in period 3. Thus, an *individual contribution* effect would invalidate the assumption of common prize spreads and explain the absence of neutrality.

**(ii) Absence of effects altering players' effort cost or ability** The second assumption underpinning neutrality is that players' cost of effort is not affected by the circumstances of

---

<sup>7</sup>A decisive – or non-trivial battle is a battle for which the winning team has not been determined yet. In a best-of-three team contest, battle 3 is decisive if and only if each team has won one battle in the two previous rounds.

the contest: the outcome of a battle only depends on players' relative abilities ( $\theta_{X_i}$  and  $\theta_{Y_i}$ ), which remain fixed.

However, players' effort cost may be affected by psychological factors related to the situation at hand. A player might have *psychological momentum* following the victory of his teammate, which would be equivalent to a decrease in his effort cost. Conversely, players' cost of effort could increase when they are under pressure. This *choking under pressure* phenomenon could occur when the stakes of the battle are particularly high for one of the players, such as the player from the lagging team involved in battle 2, whose defeat would lead to collective failure. Incorporating such effects in the cost function of players would also lead to non-neutrality.

These are the phenomena that might explain why these two crucial assumptions do not necessarily hold. The next section presents an empirical strategy to test for neutrality.

### 2.3.2 Empirical setting and data

#### International squash championships as a randomized natural experiment

Professional squash team data are particularly well suited to analysis of multiple pairwise battles. The structure of international squash competitions mirrors a theoretical best-of-three team contest with complete information: both the identity of the six players (three in each rival team) taking part and the order in which they play are determined before the beginning of the contest.<sup>8</sup> Battles are played sequentially; each player only plays one battle. A team wins as soon as two of its players win.

Before a team championship starts, each National Squash Association has to rank its players by descending order of strength and has to declare this order truthfully: opponents or organizers may object to a within-team ranking that does not reflect the actual hierarchy amongst teammates. This within-team ranking applies to all the contests disputed by the team during the championship. In each team contest, the order of the three individual battles is randomly drawn from four possibilities: 1-2-3, meaning that players ranked first play the first game, players ranked second play the second game and players ranked third play the third game, 1-3-2, 2-1-3,

---

<sup>8</sup>Given the high stakes of international team championships, only the best players from each participating country are selected to compete. Selection in the national team is based on each player's performance in the various individual championships before the team event. Performance in the national team has no bearing on players' individual ranking.

and 3-1-2. This ex-ante randomly-drawn order of play ensures that teams cannot manipulate the sequence of games to be played in any way.<sup>9</sup> Thanks to this unique feature, we can use team squash contests as a randomized natural experiment to analyze team effects in multiple pairwise battles.

## Data

We construct a comprehensive dataset of international squash team confrontations from 1998 to 2016<sup>10</sup> that includes 2,039 national team matches. We consider 55 international team tournaments, including Men’s and Women’s World Team Championships, Men’s and Women’s Asian Team Championships and Women’s European Team Championships.<sup>11</sup> The World Team Championships are organized by the World Squash Federation (WSF). The competition is held once every two years, each time in a different venue. The men’s and women’s events are held separately in different years.<sup>12</sup> The Asian Team Championships are organized by the Asian Squash Federation (ASF) and take place every two years. Finally, the European Squash Federation (ESF) holds the European Team Championships annually.

We have also recorded additional information: match durations (for most entries), locations and exact scores. The official scoring system for all levels of professional and amateur squash is called “point-a-rally scoring” (PARS). In PARS, the winner of a rally always receives a point, regardless of whether he served or returned. Sets are now played to 11, but were played to 9 until 2007 at Men’s World Team Championships, 2008 at Women’s World Team Championships, 2009 at Women’s European Team Championships, and 2010 at Men’s and Women’s Asian Team Championships. Players win a set by two clear points, i.e., if the score reaches 10–10, play continues until one player wins two consecutive points. Battles are the best-of-five sets, and the contest is a best-of-three battle.

Our data also encompass professional players’ monthly world rankings, which are published by the Professional Squash Association (PSA).<sup>13</sup> These rankings are based solely on players’ performance in individual tournaments and, as such, are not correlated with their performance

---

<sup>9</sup>See Section S1 of the WSF Regulations, Section L1 of the ESF Regulations and Section T1 of the ASF Regulations for more details.

<sup>10</sup>The data were gathered from the website <http://www.squashinfo.com>.

<sup>11</sup>We do not include Men’s European Team Championships in our sample because the tournament uses a best-of-four structure with ties broken by points count back.

<sup>12</sup>The 2015 Men’s World Team Championship, which was to be held in Cairo, Egypt, has been canceled.

<sup>13</sup>Note that amateur squash players that might be involved in team championships have no PSA ranking.

in past team tournaments. We use the PSA rankings as a proxy for players' ability.

Finally, we also collected data on teams' rankings as teams are seeded before the beginning of each championship. Teams' rankings reflect the extent to which they are favorites and are determined before the beginning of the competition by specialists, who base their judgment on all available information. As such, teams' rankings encompass most of the environment effects that may be at play, including the quality of teams' management, the cohesiveness between players and their current physical condition.

### Descriptive statistics

This section provides a series of descriptive statistics on international squash team confrontations from 1998 to 2016.

We first compare the rankings of the players involved in battle 1, battle 2 and battle 3. This allows us to ensure that (i) PSA rankings reflect correctly players' ability and (ii) the professional squash events comply with World Squash Championship Regulations. According to these regulations, each National Squash Association has to rank its players by descending order of strength. We do not identify each player's position (i.e., first, second or third) within his team, but this should be correlated with PSA rankings. Moreover, these regulations impose that the order of the three battles is randomly drawn from the following four possibilities: 1-2-3, 1-3-2, 2-1-3, and 3-1-2. Therefore, the corresponding theoretical probabilities that battle  $i$  involves a player ranked first, second or third are as follows (see table 2.1).

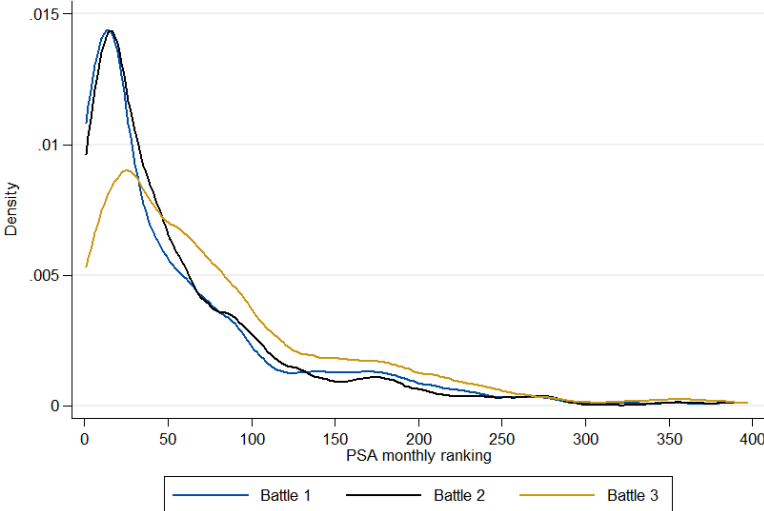
Table 2.1: Within-team rankings of players involved in each battle according to WSF regulation

	Battle 1	Battle 2	Battle 3
Probability that battle $i$ involves players ranked			
First	50%	50%	0%
Second	25%	25%	50%
Third	25%	25%	50%
Total	100%	100%	100%

Accordingly, the distribution of the PSA rankings of the players involved in battle 1 and battle 2 should be similar: the first two confrontations involve players who are ranked first with a 50% probability and ranked second or third with a 25% probability. On average, the PSA rankings of the players involved in battle 3 should be higher than that of those competing in battle 1 and battle 2. In battle 3, there are no top-ranked players, and players ranked second and

third are equally distributed. Figure 2.1 provides the distribution of the rankings of the players involved in battle 1, battle 2 and battle 3 in our dataset. As expected, ranking distributions are similar for the first two rounds and players involved in battle 3 have a higher PSA ranking than teammates involved in previous confrontations. These findings are perfectly consistent with team squash regulations and show that PSA rankings correctly reflect players' ability.

Figure 2.1: Distribution of players' monthly PSA rankings, per round – 1998-2016



Furthermore, we compare the characteristics of the matches that take place in battle 1, battle 2, and battle 3. We do so to check if there are any significant differences between these confrontations to ensure that, in battle 2, the player in the lagging team does not adopt a risky strategy, as he must win to keep his team in the contest. If this were the case, one would expect to observe significant differences amongst the main characteristics of battle 1 and battle 2 matches. In particular, risky strategies should reduce the duration of a match: a player who gambles effectively tries to shorten each rally by attempting winning shots.

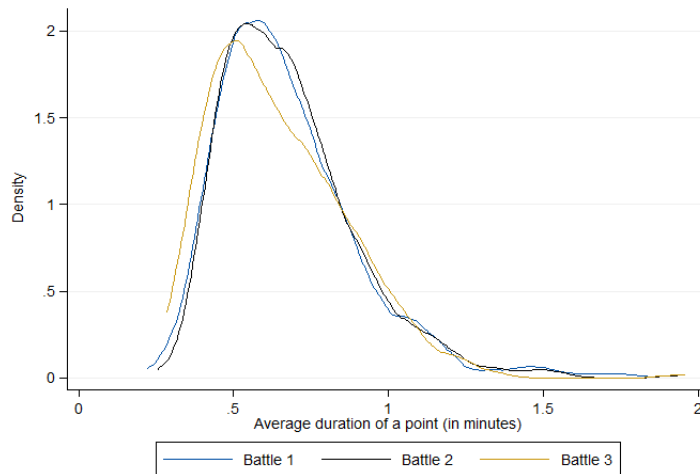
Table 2.2 displays, for each round (i.e., battle 1, battle 2 and battle 3), the average number of sets per match; the proportion of three-set matches (whose final score is necessarily 3-0), four-set matches (3-1) and five-set matches (3-2); the average number of points per match and per set; the average proportion of points won by the winner; and the average match duration.

Table 2.2: Characteristics of international squash team matches, per round – 1998-2016

	Battle 1	Battle 2	Battle 3	Total
Average number of sets per match	3.5	3.5	3.5	3.5
% of matches with a final score of				
3-0	60.7%	61.2%	64.9%	61.8%
3-1	26.2%	23.4%	23.7%	24.6%
3-2	13.1%	15.3%	11.4%	13.6%
Average number of points per match	52.2	53.6	52.8	52.9
Average number of points per set	14.9	15.3	15.1	15.1
Average duration (minutes)	38.2	38.8	33.7	37.3

These descriptive statistics show that battle 1 and battle 2 matches have similar characteristics. In addition, figure 2.2 shows the distribution of the average duration of a point in matches that take place in battle 1, battle 2 and battle 3. For each confrontation, we compute the ratio of the total duration of the match and the number of points played. If the players who belong to the lagging team gambled in battle 2, the average duration of a rally in battle 2 would be shorter than in battle 1, where none of the players has reason to adopt a particularly risky strategy. This is not what is observed in our data: Figure 2.2 shows that the distribution of the average duration of points played in battle 2 matches is the same as in battle 1 games.

Figure 2.2: Distribution of the average duration of points played in squash matches, per round – 1998-2016



The descriptive statistics confirm that players do not adopt a risky strategy based on the result of the first confrontation. Battle 1 and battle 2 matches display similar observable char-

acteristics.

### Measures of players' relative ability

In this paper, we want to estimate how past battles may affect the current battle outcome, in order to further analyze individual behaviors in teams. The outcome of a given battle mainly depends on the relative ability of both competing players. As mentioned in section 2.3.2, P-SA rankings correctly reflect players' ability. In our empirical analysis, we use three different measures of players' ability based on PSA rankings as control variables.

First, we use rankings as a categorical variable with seven modalities: Top 5; 6-15; 16-30; 31-50; 51-75; 76-105; and 106-450. These specific modalities are constructed by increasing the size of the ranking range by 5 from one category to the next (except for the last one). This accounts for the fact that a small gap between two top players reflects a significant difference in their respective ability compared to a similar ranking gap between two second-tier professional players. These categories allow us to strike a good balance between an accurate measure of players' ability and a sufficient number of observations in each modality.

Second, we introduce each possible interaction between the aforementioned categories into our empirical models (e.g. Top 5 vs 16-30). We therefore generate 49 variables, which correspond to the couple of ranking categories for each pair of competing players.

Finally, we consider the ratio of players' rankings in order to show that our results are not dependent on the choice of ranking categories. We consider the ratio of the ranking of the best player, with the lowest ranking, against the ranking of the worst, so that the ratio always lies between 0 and 1.

All the results presented in the remainder of this paper are robust to these three different measures of players' relative ability.

### 2.3.3 Testing for neutrality in multiple pairwise battles

According to Fu, Lu & Pan (2015)'s model, the probability of winning a battle is not affected by the outcome of previous battles – all that matters is the relative ability of the players involved in a given battle. Neutrality is derived from the fact that both players have the same incentive to win because they have the same prize spread (i.e., the same utility gap between winning and losing).

**Test 1.** *There is evidence in support of neutrality if winning battle 1 does not affect the probability of winning battle 2.*



## The absence of neutrality: statistical evidence

The most direct way to assess whether winning the first battle affects the probability of winning the second is to construct a sample in which players from both teams involved in the second battle have similar rankings. Based on this sample of equally skilled players, one would expect, if there were neutrality, half of the contests to be won by the player who belongs to the leading team.<sup>14</sup>

We use two different methods to restrict our sample of equally skilled players. The first method is based on the ratio of the rankings of both players involved in the second battle, and we restrict our sample to observations where this ratio is lower than certain thresholds: (i)  $ratio < 1.5$  – variant 1, (ii)  $ratio < 1.4$  – variant 2, (iii)  $ratio < 1.3$  – variant 3 and finally (iv)  $ratio < 1.2$  – variant 4. According to this definition, a match between a player ranked 15 and a player ranked 25 will not be included in any variant (the ratio of these rankings being 1.66), while a match between a player ranked 15 and a player ranked 17 will be included in the four variants (the ratio of these rankings being 1.13). The second approach consists in considering only battle 2 matches in which both players' rankings belong to the same category (Top 5; 6-15; 16-30; 31-5; 51-75; 76-105; and 106-450). Therefore, we only take into account a confrontation that involves a player ranked e.g. 6-15 if his opponent's ranking lies within the same category.

We note  $X_1$  the player who won the first battle against  $Y_1$ , and  $X_2$  the player who belongs to the leading team involved in battle 2 against  $Y_2$ . Table 2.3 reports the empirical probability that  $X_2$  wins the second battle for each of the variants considered.

Table 2.3: Statistical evidence against neutrality

	Ratio of rankings lower than				Same category
	1.5	1.4	1.3	1.2	
$X_2$ wins battle 2	<b>59.7%***</b>	<b>59.1%**</b>	<b>60.4%**</b>	<b>60.4%**</b>	<b>56.7%*</b>
Number of observations	211	181	139	91	203

Statistically different from 50% at \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

The results presented in table 2.3 show that the probability that the player on the leading team wins is greater than 50%. In other terms, the figures suggest the absence of neutrality: winning the first battle increases the probability of winning the second one.

<sup>14</sup>This type of identification strategy is implemented by Malueg & Yates (2010), who construct a sample of tennis matches with equally skilled players.

## Evidence against neutrality: main specification

Restricting the sample to players who have similar rankings is a simple way to control for players' relative ability but it considerably reduces the number of observations. We therefore integrate our measures of players' relative ability as a control variable in order to use our entire sample.

To do so, we label the two opposing teams as "Team A" and "Team B"<sup>15</sup> and their players as  $A_1, A_2, A_3, B_1, B_2$  and  $B_3$ , where the subscript indicates the battle in which the player is engaged. We can test for neutrality by assessing whether the probability that  $A_2$  defeats  $B_2$  is higher when  $A_1$  won against  $B_1$  in the previous battle, controlling for  $A_2$ 's and  $B_2$ 's ability. Thus we regress the dummy variable indicating whether  $A_2$  wins or loses battle 2 on a dummy variable indicating whether  $A_1$  won or lost battle 1 and on a measure of  $A_2$  and  $B_2$ 's relative ability based on their rankings. The tested econometric specification is as follows.

$$A_2 \text{ wins battle 2} = \beta_0 + \beta_{\text{Non-neutrality}} \times A_1 \text{ won battle 1} + f(\text{Ranking}_{A_2}, \text{Ranking}_{B_2}) + \epsilon_{AB2},$$

where  $f(\text{Ranking}_{A_2}, \text{Ranking}_{B_2})$  refers to one of the three measures of players' relative ability described in the previous section: i) ranking modalities, (ii) interacted ranking modalities, (iii) the ratio of rankings. We integrate these measures of players' ability taking into account the symmetric structure of the dataset: the outcome of a battle depends on the characteristics of the two players. Hence, when we use ranking modalities and interacted modalities, we break down every battle into two observations and weight each observation by  $\frac{1}{2}$  so as to adjust standard errors correctly. When we use the ratio of rankings as a proxy for relative ability, we define player  $A_i$  as the player who has the better ranking, so that the ratio of rankings always lies between 0 and 1.

We use a linear probability model as it makes the interpretation of the coefficients of interest easy.<sup>16</sup> The results are displayed in columns (1), (3), (5) of table 2.4. The coefficient associated with  $A_1$  won battle 1 is significant at the 0.1% level using any of the measures of players' ability. The magnitude of the effect is very strong (from 9.8 to 14.2 percentage points depending on the ability measure used). This means that in a battle involving two players with similar rankings, the player on the leading team wins with a probability that lies between 54.9% and 57.1%.

---

<sup>15</sup>In the remainder of this paper, we label "Team A" and "Team B" each of the opposing teams in a given confrontation, with no further conditions on the outcome of the first battle. When we deliberately choose the team that won the first battle, we refer to it as "Team X", or "X".

<sup>16</sup>We obtain very similar results when we use a probit estimation (see tables 2.12 and 2.13 in Appendix B).

This series of results suggests the absence of neutrality. However, as explained before, we need to account for potential peer effects. Being on a team with high-performing teammates may increase a player's productivity, as a more stimulating environment may boost performance. Since high-performing players tend to win their battle, the player on the leading team is likely to be surrounded by more talented teammates than the player on the lagging team. Therefore, peer effects might be a confounding factor for sequence dependence. We take into account environment effects and other unobservables, such as the relative quality of the teams' managers and the cohesiveness between players, by including the teams' rankings as additional continuous control variables in specifications (2), (4) and (6). We also add controls regarding the location of the confrontation, as playing at home can affect the outcome. Accordingly, we add dummies indicating whether team A is playing at home or away (the reference being the neutral field).

Winning the first battle remains significant at the 0.1% level once teams' rankings are introduced. The magnitude of the effect decreases slightly but remains substantial (from 8.7 to 11.4 percentage points depending on the specification). This is clear evidence that sequence dependence is not caused by confounding peer effects.

Table 2.4: Evidence against neutrality

<i>Dep. var.</i> : $A_2$ wins battle 2	(1)	(2)	(3)	(4)	(5)	(6)
<b><math>A_1</math> won battle 1</b>	0.142*** (0.029)	0.114*** (0.031)	0.128*** (0.029)	0.097** (0.031)	0.098*** (0.028)	0.087** (0.030)
$A_2$ 's ranking: Top 5	0.721*** (0.057)	0.646*** (0.070)				
$A_2$ 's ranking: 6-15	0.576*** (0.046)	0.518*** (0.056)				
$A_2$ 's ranking: 16-30	0.466*** (0.043)	0.417*** (0.050)				
$A_2$ 's ranking: 31-50	0.293*** (0.043)	0.260*** (0.047)				
$A_2$ 's ranking: 51-75	0.184*** (0.046)	0.154** (0.049)				
$A_2$ 's ranking: 76-105	0.093* (0.047)	0.073 (0.049)				
$B_2$ 's ranking: Top 5	-0.721*** (0.057)	-0.646*** (0.070)				
$B_2$ 's ranking: 6-15	-0.576*** (0.046)	-0.518*** (0.056)				
$B_2$ 's ranking: 16-30	-0.466*** (0.043)	-0.417*** (0.050)				
$B_2$ 's ranking: 31-50	-0.293*** (0.043)	-0.260*** (0.047)				
$B_2$ 's ranking: 51-75	-0.184*** (0.046)	-0.154** (0.049)				
$B_2$ 's ranking: 76-105	-0.093* (0.047)	-0.073 (0.049)				
$A_2$ 's ranking vs $B_2$ 's ranking			X	X		
$\frac{Ranking_{A_2}}{Ranking_{B_2}} (< 1)$					-0.535*** (0.046)	-0.522*** (0.054)
$A_2$ at home		0.009 (0.052)		0.010 (0.052)		0.085 (0.049)
$B_2$ at home		-0.009 (0.052)		-0.010 (0.052)		0.048 (0.050)
$A_2$ 's team ranking		-0.007* (0.004)		-0.008* (0.004)		-0.008* (0.004)
$B_2$ 's team ranking		0.007* (0.004)		0.008* (0.004)		0.004 (0.003)
Constant	0.429*** (0.037)	0.443*** (0.058)	0.436*** (0.048)	0.451*** (0.069)	0.436*** (0.048)	0.451*** (0.069)
<b>Controls for players' ability</b>						
Ranking category	YES	YES	NO	NO	NO	NO
Interaction of rkg categories	NO	NO	YES	YES	NO	NO
Ratio of rankings	NO	NO	NO	NO	YES	YES
Observations	934	896	934	896	934	896
$R^2$	0.42	0.42	0.45	0.46	0.17	0.17

Standard errors in parentheses: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Reading note (column 1): the probability that  $A_2$  wins increases by 0.142 when his teammate  $A_1$  won battle 1.

Reading note (column 1): the probability of winning of a player ranked in the Top 5 is 0.721 higher than the probability of winning of a player ranked 106-450 (reference category).

Note: the number of observations decreases from 934 in specifications (1), (3), (5) to 896 in specifications (2), (4),(6) because there are 38 contests for which teams' rankings are missing.

We find strong evidence that winning the first battle increases the probability of winning the second one. This contradicts neutrality.

## 2.4 The role of individual contribution to team success

In this section, we explain the absence of neutrality by providing compelling evidence, based on the predictions of the model developed in section 2.3.1, that individuals are willing to contribute to the success of their team. The outcomes of subsequent battles are dependent because players value being responsible for their team’s success. Outcome dependence is driven by what we refer to as “individual contribution.” Furthermore, we perform robustness checks, which confirm the existence of individual contribution in multiple pairwise battles and rule out alternative explanations developed in the existing literature on other team settings.

### 2.4.1 Evidence for individual contribution

Players might value being partly responsible for the success of their team. If players individually value their contribution to the team, they get an additional reward  $c$  ( $c > 0$ ) when their victory leads their team to success. Table 2.5 displays players’ payoffs when there is individual contribution.

Table 2.5: Payoffs in the individual contribution scenario

	Player’s team wins	Player’s team loses
Player wins	$v + V + c$	$v$
Player loses	$V$	$0$

In such a case, the main intuition is that the player on the leading team would have more incentive to win the second battle than the player on the lagging team because the former is sure to contribute to his team’s success if he wins while the player on the lagging team will be “success-responsible” if and only if his teammate also wins the third battle. This asymmetry of incentives between the two players depends on the expected outcome of battle 3. For example, if  $X_1$  wins battle 1 and  $X_3$  has a extremely low probability of winning the third match (i.e.,  $p_3^*$  falls to 0), both players can contribute to their team’s victory by winning battle 2, and both players would make a symmetrical positive effort to get the additional reward. In this extreme case, winning the first battle should have no effect on the probability of winning the second

one. Conversely, in the extreme case where  $X_1$  wins battle 1 and  $X_3$  has an extremely high probability of winning the third match (i.e.,  $p_3^*$  increases to 1), the asymmetry between the two players reaches its maximum:  $X_2$  is certain to receive the contribution reward if he wins while  $Y_2$  has no chance of getting it.

**Main test** Formally, we obtain the following predictions, which confirm the role played by  $p_3^*$  in the individual contribution scenario:<sup>17</sup>

$$p_{2IC}^* = \frac{\theta_{X_2}(v + (1 - p_3^*)V + c)}{\theta_{X_2}(v + (1 - p_3^*)V + c) + \theta_{Y_2}(v + (1 - p_3^*)(V + c))},$$

where team  $X$  won battle 1.

This gives the two following results: first,  $p_{2IC}^* > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$ , so winning battle 1 increases the probability of winning battle 2. Second,  $\frac{\partial p_{2IC}^*}{\partial p_3^*} > 0$  and  $\frac{\partial(1 - p_{2IC}^*)}{\partial(1 - p_3^*)} > 0$ , so the probability of winning battle 2 increases with the teammate's probability of winning battle 3. This allows us to derive the following empirical test for individual contribution.

**Test 2.** *There is evidence in support of an “individual contribution” effect if*

- *Winning battle 1 increases the probability of winning battle 2.*
- *The probability of winning battle 2 increases with the probability of winning battle 3.*

The first condition given by test 2 is satisfied, as the results of test 1 show. We can test for the second condition of test 2 by assessing whether the probability that  $A_2$  wins against  $B_2$  increases with the extent to which  $A_3$  is the favorite in battle 3. We use the gap between  $A_3$ 's and  $B_3$ 's ranking modalities, labelled  $RM_{A_3} - RM_{B_3}$ , as a proxy for the probability of winning of  $A_3$ . For example, when  $A_3$  ranks in the Top 5 and  $B_3$  ranks between 16 and 30,  $RM_{A_3} - RM_{B_3} = 2$ .<sup>18</sup>

We regress the dummy variable indicating whether  $A_2$  wins battle 2 on  $RM_{A_3} - RM_{B_3}$ , on a measure of  $A_2$ 's and  $B_2$ 's relative ability and on the control variables used previously (playing home/away and teams' rankings). The tested econometric specification is therefore as follows.

$$A_2 \text{ wins battle 2} = \beta_0 + \beta_{IC} \times (RM_{A_3} - RM_{B_3}) + f(Ranking_{A_2}, Ranking_{B_2}) \\ + \beta_{home} \times Home_A + \beta_{away} \times Away_A + \beta_{tA} \times Team's ranking_A - \beta_{tB} \times Team's ranking_B + \epsilon_{AB2}.$$

---

<sup>17</sup>See Appendix for detailed computations.

<sup>18</sup>In order to perform the estimation on the same sample as the one used to test for neutrality (table 2.4), we create an additional ranking modality for amateur squash players involved in battle 3, whose level is too low to have a PSA ranking. Hence, if a player  $A_3$  ranked 51-75 is opposed to an amateur player  $B_3$ ,  $RM_{A_3} - RM_{B_3} = 2$ .

The results obtained with the three measures of players' relative ability are reported in columns (1), (3) and (5) of table 2.6. As predicted in the individual contribution scenario, the variable  $RM_{A_3} - RM_{B_3}$  is positive and significant at the 1% level in the three specifications, showing that the probability of winning battle 2 increases with the probability of winning battle 3. When the difference in ranking modalities between the two players involved in battle 3 increases by 1, the probability of winning battle 2 increases by about 2 percentage points. Hence, a gap of 5 ranking modalities – which comes very close to the extreme case where  $p_3^*$  goes to 1 – increases the probability of winning battle 2 by 0.1. This effect is about as strong as the estimated effect of winning battle 1 (see table 2.4). This finding is perfectly consistent with the individual contribution effect, according to which winning battle 1 has no effect on battle 2 when the opposing team is expected to win battle 3.

One potential concern with specifications (1), (3) and (5) is confounding peer effects: being the favorite in battle 3 might be significant because it might imply that the player is in a more stimulating environment with more able teammates. If such an effect were at play, being the favorite in battle 1 should have the same effect, as there is no reason to believe that the influence of the teammate involved in battle 1 would be different from the influence of the teammate playing battle 3. In specifications (2), (4) and (6), we include the gap between  $A_1$ 's and  $B_1$ 's ranking modalities (labelled  $RM_{A_1} - RM_{B_1}$ ) as a control to test for peer effects. The variable  $RM_{A_1} - RM_{B_1}$  is not significant in any of the specifications and its inclusion does not affect the coefficient associated with our variable of interest,  $RM_{A_3} - RM_{B_3}$ . This confirms that peer effects are not at play and provides compelling evidence that individual contribution drives the observed linkage between subsequent battles.

Table 2.6: Individual contribution – Main test

<i>Dep. var.</i> : $A_2$ wins battle 2	(1)	(2)	(3)	(4)	(5)	(6)
$RM_{A_3} - RM_{B_3}$	0.024*** (0.007)	0.024*** (0.007)	0.023** (0.007)	0.024** (0.007)	0.021** (0.007)	0.021** (0.007)
$RM_{A_1} - RM_{B_1}$		-0.002 (0.007)		-0.004 (0.007)		-0.004 (0.007)
$A_2$ 's ranking: Top 5	0.625*** (0.072)	0.630*** (0.073)				
$A_2$ 's ranking: 6-15	0.513*** (0.058)	0.516*** (0.059)				
$A_2$ 's ranking: 16-30	0.424*** (0.051)	0.426*** (0.052)				
$A_2$ 's ranking: 31-50	0.273*** (0.047)	0.275*** (0.047)				
$A_2$ 's ranking: 51-75	0.174*** (0.050)	0.175*** (0.050)				
$A_2$ 's ranking: 76-105	0.082 (0.050)	0.083 (0.050)				
$B_2$ 's ranking: Top 5	-0.625*** (0.072)	-0.630*** (0.073)				
$B_2$ 's ranking: 6-15	-0.513*** (0.058)	-0.516*** (0.059)				
$B_2$ 's ranking: 16-30	-0.424*** (0.051)	-0.426*** (0.052)				
$B_2$ 's ranking: 31-50	-0.273*** (0.047)	-0.275*** (0.047)				
$B_2$ 's ranking: 51-75	-0.174*** (0.050)	-0.175*** (0.050)				
$B_2$ 's ranking: 76-105	-0.082 (0.050)	-0.083 (0.050)				
$A_2$ 's ranking vs $B_2$ 's ranking			X	X		
$\frac{Ranking_{A_2}}{Ranking_{B_2}} (< 1)$					-0.503*** (0.055)	-0.508*** (0.056)
$A_2$ at home	0.009 (0.052)	0.009 (0.053)	0.010 (0.052)	0.010 (0.052)	0.076 (0.048)	0.075 (0.048)
$B_2$ at home	-0.009 (0.052)	-0.009 (0.053)	-0.010 (0.052)	-0.010 (0.052)	0.045 (0.050)	0.043 (0.050)
$A_2$ 's team ranking	-0.006 (0.004)	-0.006 (0.004)	-0.006 (0.004)	-0.007 (0.004)	0.002 (0.004)	0.001 (0.004)
$B_2$ 's team ranking	0.006 (0.004)	0.006 (0.004)	0.006 (0.004)	0.007 (0.004)	0.002 (0.003)	0.003 (0.003)
Constant	0.500*** (0.057)	0.500*** (0.057)	0.500*** (0.069)	0.500*** (0.069)	0.946*** (0.038)	0.952*** (0.040)
<b>Controls for players' ability</b>						
Ranking category	YES	YES	NO	NO	NO	NO
Interaction of rkg categories	NO	NO	YES	YES	NO	NO
Ratio of rankings	NO	NO	NO	NO	YES	YES
Observations	874	874	874	874	874	874
$R^2$	0.43	0.43	0.47	0.47	0.17	0.17

Standard errors in parentheses.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Note: the number of observations drops from 896 in table 2.4 to 874 in this table as 22 trivial battles 3 are not played.



**Additional test** One may argue that the individual contribution effect should also be at play in a trivial battle 3 – where, by definition, the winning team has already been determined. In a best-of-three team contest, battle 3 is trivial if one team has already won the two previous confrontations. A player involved in a trivial battle 3 whose team has already won the contest should also value winning his match, so as to take part, albeit less directly than his teammates, in the success of his team. Winning allows him to be part of his team’s victory even if he was randomly assigned to the third battle, which happens to be trivial. Therefore, if individual contribution is at play, we should observe that his probability of winning the last trivial battle is *ceteris paribus* higher than his opponent’s. Formally, we obtain the following prediction:<sup>19</sup>

$$p_{3ICtrivial}^* = \frac{\theta_{X_3}(v+c)}{\theta_{X_3}(v+c) + \theta_{Y_3}v},$$

where team  $X$  won battle 1 and battle 2. As  $p_{3ICtrivial}^* > \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}$ , we can derive the following empirical test for individual contribution.

**Test 3.** *There is evidence in support of an “individual contribution” effect if winning battle 1 and battle 2 increases the probability of winning trivial battle 3.*

We implement this additional test by restricting our sample to trivial battles 3<sup>20</sup> and assessing whether being in the team that won the first two battles increases the probability of winning the last. The tested econometric specification is as follows.

$$A_3 \text{ wins battle 3} = \beta_0 + \beta_{IC} \times A_1 \text{ won battle 1 and } A_2 \text{ won battle 2} + f(\text{Ranking}_{A_3}, \text{Ranking}_{B_3}) + \beta_{home} \times Home_A + \beta_{away} \times Away_A + \beta_{tA} \times Team's \text{ ranking}_A - \beta_{tB} \times Team's \text{ ranking}_B + \epsilon_{AB3}.$$

The results are reported in table 2.7 for the three measures of the relative ability of players involved in trivial battles 3. The coefficient of interest is statistically significant and positive in the three specifications. This confirms that a player – whatever the stake of the battle he has been randomly assigned to – is motivated to participate in his team’s success.

---

<sup>19</sup>See Appendix for detailed computations.

<sup>20</sup>This is why the number of observations drops to 378 in table 2.7.

Table 2.7: Individual contribution – Additional test

<i>Dep. var:</i> $A_3$ wins battle 3	(1)	(2)	(3)
$A_1$ won battle 1 and $A_2$ won battle 2	0.273*** (0.055)	0.176** (0.060)	0.133* (0.055)
$A_3$ 's ranking: Top 5	0.417*** (0.111)		
$A_3$ 's ranking: 6-15	0.326*** (0.077)		
$A_3$ 's ranking: 16-30	0.318*** (0.068)		
$A_3$ 's ranking: 31-50	0.191** (0.062)		
$A_3$ 's ranking: 51-75	0.059 (0.061)		
$A_3$ 's ranking: 76-105	0.122* (0.058)		
$B_3$ 's ranking: Top 5	-0.417*** (0.111)		
$B_3$ 's ranking: 6-15	-0.326*** (0.077)		
$B_3$ 's ranking: 16-30	-0.318*** (0.068)		
$B_3$ 's ranking: 31-50	-0.191** (0.062)		
$B_3$ 's ranking: 51-75	-0.059 (0.061)		
$B_3$ 's ranking: 76-105	-0.122* (0.058)		
$A_3$ 's ranking vs $B_3$ 's ranking		X	
$\frac{Ranking_{A_3}}{Ranking_{B_3}} (< 1)$			-0.241** (0.076)
$A_3$ at home	0.064 (0.071)	0.042 (0.072)	0.124* (0.058)
$B_3$ at home	-0.064 (0.071)	-0.042 (0.072)	0.075 (0.079)
$A_3$ 's team ranking	-0.012** (0.004)	-0.014** (0.005)	-0.012* (0.005)
$B_3$ 's team ranking	0.012** (0.004)	0.014** (0.005)	0.008* (0.004)
Constant	0.364*** (0.075)	0.412*** (0.088)	0.823*** (0.068)
<b>Controls for players' ability</b>			
Ranking category	YES	NO	NO
Interaction of rkg categories	NO	YES	NO
Ratio of rankings	NO	NO	YES
Observations	378	378	378
$R^2$	0.59	0.64	0.17

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 

## 2.4.2 Robustness checks and alternative explanations

The empirical strategy based on our theoretical predictions shows that the dynamic linkage between subsequent battles observed in our data is in line with the individual contribution effect, according to which players value being responsible for their team's success.

We confirm the existence of individual contribution by testing for alternative effects, which could also explain the observed dependence of the outcomes of the first two battles. The theoretical model presented in section 2.3.1 shows that the dynamic linkage observed between two subsequent battles can either result from effects that endogenously generate a gap in players'

prize spreads, or from effects that alter players' ability.

These effects are summarized in table 2.8.

Table 2.8: Individual contribution and alternative explanations

Type	Effect	Description
Asymmetry in prize spreads	Individual contribution	Individuals value being responsible for collective success
	Guilt aversion	Individuals dread being responsible for collective failure
	Altruism	Individuals internalize their teammates' costs of effort
Alteration of players' ability	Choking under pressure	Pressure has a detrimental effect on performance
	Momentum	Recent success increases confidence

### Other effects that generate a gap in prize spreads: Guilt aversion and altruism

**Guilt aversion** Players might suffer from being (partly) responsible for the failure of their team. Charness & Dufwenberg (2006) provide experimental evidence consistent with people striving to live up to others' expectations so as to avoid guilt. Chen & Lim (2013) analyze whether managers should organize employees to compete in teams or as individuals. Their main conclusion according to which team-based contests yield greater effort than individual-based contests, is rooted in contestants' aversion to letting their team down.

In our setting, guilt aversion implies that a player who loses his battle bears an additional loss ( $-s$ ,  $s > 0$ ) if his team loses the contest. This additional loss asymmetrically affects players' prize spreads and therefore may explain the absence of neutrality (see table 2.9).

Table 2.9: Payoffs in the guilt aversion scenario

	Player's team wins	Player's team loses
Player wins	$v + V$	$v$
Player loses	$V$	$-s$

Under this scenario, we derive the following predictions:<sup>21</sup>

$$p_{2GA}^* = \frac{\theta_{X_2}(v + (1 - p_3^*)(V + s))}{\theta_{X_2}(v + (1 - p_3^*)(V + s)) + \theta_{Y_2}(v + (1 - p_3^*)V + s)},$$

where team  $X$  is defined as the team that won battle 1.  $p_{2GA}^* < \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$ , which yields the following empirical test.

<sup>21</sup>See Appendix for detailed computations.

**Test 4.** *There is evidence in support of guilt aversion if winning battle 1 decreases the probability of winning battle 2.*

In our setting, winning battle 1 increases the probability of winning battle 2 (see table 2.4). Non-neutrality is not driven by guilt aversion.

**Altruism** Our results up to this point have been based on the assumption that players do not take into account their teammates' costs of effort. However, the observed link between the first two battles of the contest could be driven by the fact that individuals internalize their teammates' effort costs. This effect could be referred to as "altruism," signifying that each player on a team maximizes his utility function, taking into account not only his own effort cost but also his teammates'. For instance, the player on the leading team involved in battle 2 could make an additional effort in order to win, thereby preventing his teammate from playing a decisive battle 3 and incurring the corresponding effort cost.

The idea that individuals internalize their teammates' costs calls for the following two remarks. First, we refer to this phenomenon as "altruism," but one may argue that a player might internalize the effort cost of teammates involved in battle 3 in order to prevent them from playing a high-stake match because it is in his interest to preserve their stamina for the next rounds. This interpretation would not affect the empirical test presented below. Second, players could internalize not only their partners' effort costs, but also the various benefits they derive from playing a match. If the benefits from playing are greater than the afferent costs, an altruistic player would behave in a way that enables his teammate to play. This would imply that winning battle 1 has a negative impact on the probability of winning battle 2, which is not supported by our data (see table 2.4).

We develop a test to address the fact that individuals may internalize their teammates' costs. This test allows us to distinguish between individual contribution and altruism. Intuitively, our identification strategy is based on the fact that, in a best-of-three contest, altruistic players involved in the first battle cannot prevent their teammates from playing a high-stake second match, and can only internalize the cost of effort of the players involved in the third (potentially trivial) battle. Accordingly, we focus on battle 1 and limit our sample to contests where the favorites in battles 2 and 3<sup>22</sup> do not belong to the same team. Hence, there are only two possible scenarios regarding future battles: either i)  $A_2$  is the favorite and  $A_3$  is the underdog or ii)  $A_2$  is

---

<sup>22</sup>A player is defined as the favorite when he has a better ranking than his opponent.

*the underdog and A<sub>3</sub> is the favorite.* According to the individual contribution effect, these two scenarios are equivalent, as player A<sub>1</sub> is equally likely to get the contribution reward in either setting. On the contrary, if players were altruistic, the scenario *A<sub>2</sub> favorite, A<sub>3</sub> underdog* would be much more motivating for player A<sub>1</sub>. Indeed, when A<sub>2</sub> is favorite, A<sub>1</sub> knows that winning battle 1 implies that his teammate A<sub>3</sub> will probably not have to play a decisive battle 3 and thereby make a significant effort. On the other hand, when A<sub>2</sub> is underdog, A<sub>1</sub> knows that winning battle 1 implies that his teammate A<sub>3</sub> is very likely to play a decisive battle 3 and to exert a significant effort. Thus, altruism implies that A<sub>1</sub> has more incentive to win when A<sub>2</sub> is the favorite and A<sub>3</sub> is the underdog than in the symmetric situation.<sup>23</sup>

**Test 5.** *There is evidence in support of*

- *Altruism if the probability that A<sub>1</sub> wins battle 1 is larger ceteris paribus in the scenario “A<sub>2</sub> favorite, A<sub>3</sub> underdog” than in the scenario “A<sub>2</sub> underdog, A<sub>3</sub> favorite.”*
- *Individual contribution if the probability that A<sub>1</sub> wins battle 1 is the same in both scenarii.*

Accordingly, we regress A<sub>1</sub>'s victory on a dummy variable indicating the situation regarding battles 2 and 3 (which equals 1 when A<sub>2</sub> is favorite and A<sub>3</sub> is underdog, and 0 when A<sub>2</sub> is the underdog and A<sub>3</sub> is the favorite), on a measure of A<sub>1</sub>'s and B<sub>1</sub>'s relative ability as well as the usual control variables (playing home/away and teams' rankings).

$$A_1 \text{ wins battle 1} = \beta_0 + \beta_{\text{altruism}} \times A_2 \text{ favorite, } A_3 \text{ underdog} + f(\text{Ranking}_{A_1}, \text{Ranking}_{B_1}) \\ + \beta_{\text{home}} \times \text{Home}_A + \beta_{\text{away}} \times \text{Away}_A + \beta_{tA} \times \text{Team's ranking}_A - \beta_{tB} \times \text{Team's ranking}_B + \epsilon_{AB1}.$$

Individual contribution predicts that the variable *A<sub>2</sub> favorite, A<sub>3</sub> underdog* has no significant effect on the probability that A<sub>1</sub> wins battle 1. On the contrary, if players were altruistic, our model would predict that the variable *A<sub>2</sub> favorite, A<sub>3</sub> underdog* should have a statistically significant positive on the probability that A<sub>1</sub> wins battle 1. Results are reported in table 2.10. The coefficient associated with the variable of interest is negative and not statistically significant using any of the three measures of players' ability. This confirms the individual contribution effect and rules out altruism: players do not internalize their teammates' effort costs in our setting.

---

<sup>23</sup>See Appendix for more formal details on this test.

Table 2.10: Evidence against altruism

<i>Dep. var:</i> $A_1$ wins battle 1	(1)	(2)	(3)
$A_2$ favorite, $A_3$ underdog	-0.086 (0.065)	-0.086 (0.071)	-0.106 (0.061)
$A_1$ 's ranking: Top 5	0.837*** (0.188)		
$A_1$ 's ranking: 6-15	0.695*** (0.153)		
$A_1$ 's ranking: 16-30	0.462*** (0.134)		
$A_1$ 's ranking: 31-50	0.361* (0.141)		
$A_1$ 's ranking: 51-75	0.301* (0.124)		
$A_1$ 's ranking: 76-105	0.167 (0.128)		
$B_1$ 's ranking: Top 5	-0.837*** (0.188)		
$B_1$ 's ranking: 6-15	-0.695*** (0.153)		
$B_1$ 's ranking: 16-30	-0.462*** (0.134)		
$B_1$ 's ranking: 31-50	-0.361* (0.141)		
$B_1$ 's ranking: 51-75	-0.301* (0.124)		
$B_1$ 's ranking: 76-105	-0.167 (0.128)		
$A_1$ 's ranking vs $B_1$ 's ranking		X	
$\frac{Ranking_{A_1}}{Ranking_{B_1}} (< 1)$			-0.522*** (0.129)
$A_1$ at home	0.080 (0.130)	0.082 (0.144)	0.137 (0.125)
$B_1$ at home	-0.080 (0.130)	-0.082 (0.144)	0.002 (0.121)
$A_1$ 's team ranking	-0.007 (0.010)	-0.005 (0.011)	-0.001 (0.010)
$B_1$ 's team ranking	0.007 (0.010)	0.005 (0.011)	0.002 (0.009)
Constant	0.543*** (0.159)	0.543** (0.189)	1.027*** (0.102)
<b>Controls for players' ability</b>			
Ranking category	YES	NO	NO
Interaction of rkg categories	NO	YES	NO
Ratio of rankings	NO	NO	YES
Observations	208	208	208
$R^2$	0.25	0.28	0.11

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 

### Effects that alter players' ability: Choking under pressure and psychological momentum

**Choking under pressure** Dynamic competitive settings may exert psychological pressure on competitors, thereby affecting their performance. The player who belongs to the lagging team might – all other things being equal – face more pressure than the player in the leading team, as the former needs to win to ensure that his team remains in the contest. Such a phenomenon might explain why we observe a positive effect of a win in the first game on the probability of

winning the next. Apestegua & Palacios-Huerta (2010) use the random nature of the order of soccer penalty shoot-outs to provide evidence of such psychological pressure. Teams that take the first kick in the sequence win the penalty shoot-out 60.5 percent of the time. Taking into account the characteristics of the setting, they attribute this significant difference in performance to psychological effects resulting from the consequences of the kicking order.<sup>24</sup> Ariely, Gneezy, Loewenstein & Mazar (2009) show, based on experimental data, that choking under pressure has a stronger deterrent effect on performance when the task is publicly observable.

We incorporate choking under pressure into our theoretical setting by multiplying by  $\eta$  ( $0 < \eta < 1$ ) the ability of players who must win their individual battle to ensure that their team remains in the contest. In  $T=2$ , the player in the leading team ( $X_2$ ) does not face pressure because losing his battle does not imply that his team loses. On the contrary, the player in the lagging team ( $Y_2$ ) might choke under pressure because losing his battle induces that his team loses. Hence, the choking-under-pressure effect yields the following prediction in battle 2:<sup>25</sup>

$$p_{2CUP}^* = \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}\eta} > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$$

where team  $X$  is defined as the team that won battle 1.

In a trivial battle 3 (*i.e.* one team has already won two battles), none of the players face pressure because the outcome of their individual battle does not affect the collective outcome. Therefore, we obtain the following prediction:<sup>26</sup>

$$p_{3CUPtrivial}^* = p_3^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}$$

where team  $X$  is defined as the team that won battles 1 and 2.

As  $p_{2CUP}^* > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$  and  $p_{3CUPtrivial}^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}$ , the following empirical test can be derived from our theoretical setting.

**Test 6.** *There is evidence in support of a “choking-under-pressure” effect if:*

1. *Winning battle 1 increases the probability of winning battle 2.*
2. *Winning battle 1 and battle 2 does not affect the probability of winning a trivial battle 3.*

---

<sup>24</sup>Kocher, Lenz & Sutter (2012) find different results using a larger sample of penalty shoot-outs.

<sup>25</sup>See Appendix for detailed computations.

<sup>26</sup>See Appendix for detailed computations.

The first condition is met (see table 2.4) while the second condition is not. Indeed, our additional test on the individual contribution effect shows that winning battles 1 and 2 has a statistically significant effect on the probability of winning a trivial battle 3. Depending on the measure of relative ability that is used, the effect ranges from 13 to 27 percentage points (see table 2.7). This allows us to reject choking under pressure and to put forward the individual contribution effect.

**Psychological momentum** Psychological momentum implies that winning a battle increases a player’s confidence and makes him more likely to win the next one (“success breeds success”). Rosenqvist & Skans (2015) provide field evidence of this phenomenon. They use data from European golf tournaments to show that players who (marginally) succeeded in making the cut substantially increased their performance in subsequent tournaments relative to players who (marginally) failed to make the cut.

Psychological momentum might also occur at the team level in multiple pairwise battles. We incorporate this effect into our theoretical setting by multiplying by  $\psi$  ( $\psi > 1$ ) the ability of the player whose team won the last battle. This changes the probability that player  $X_2$  and player  $X_3$  win. In this case,<sup>27</sup>

$$p_{3PM}^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}\psi},$$

where team  $X$  is defined as the team that won battle 1 and lost battle 2.

$$p_{2PM}^* = \frac{\theta_{X_2}\psi}{\theta_{X_2}\psi + \theta_{Y_2}},$$

where team  $X$  is defined as the team that won battle 1.

As  $p_{2PM}^* > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$  and  $p_{3PM}^* < \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}$ , the following empirical test can be derived from our theoretical setting.

**Test 7.** *There is evidence in support of psychological momentum if:*

1. *Winning battle 1 increases the probability of winning battle 2.*
2. *In a non-trivial battle 3, the player in the team that won battle 2 is more likely to win than the player in the team that won battle 1.*<sup>28</sup>

---

<sup>27</sup>See Appendix for detailed computations.

<sup>28</sup>This identification strategy is also used by Malueg & Yates (2010) and Mago et al. (2013).



To test for the second condition, we focus on the subsample of non-trivial battle 3s (i.e., the matches in which the winning team had not been determined after the first two battles). For these matches, there are only two possible scenarios regarding the outcome of the two previous battles: either  $A_1$  won battle 1 and  $A_2$  lost battle 2, or  $A_1$  lost battle 1 and  $A_2$  won battle 2. We create a dummy variable that is equal to 0 in the first scenario and to 1 in the second scenario. Psychological momentum would imply that this variable has a positive and statistically significant effect on  $A_3$  wins battle 3. On the contrary, individual contribution predicts that this variable should not have any effect, as both players face symmetric incentives in a decisive battle 3. The model is therefore given by:

$$A_3 \text{ wins battle 3} = \beta_0 + \beta_{PM} \times A_1 \text{ lost battle 1 and } A_2 \text{ won battle 2} + f(\text{Ranking}_{A_3}, \text{Ranking}_{B_3}) \\ + \beta_{home} \times Home_A + \beta_{away} \times Away_A + \beta_{tA} \times Team's \text{ ranking}_A - \beta_{tB} \times Team's \text{ ranking}_B + \epsilon_{AB3}$$

Table 2.11 reports the results obtained with the three measures of players' ability. The effect of the sequence variable  $A_1$  lost battle 1 and  $A_2$  won battle 2 is not statistically significant in any of the three specifications. This is evidence in favor of individual contribution and against psychological momentum.

Table 2.11: Evidence against psychological momentum

<i>Dep. var:</i> $A_3$ wins battle 3	(1)	(2)	(3)
$A_1$ lost battle 1 and $A_2$ won battle 2	-0.026 (0.067)	-0.027 (0.072)	-0.002 (0.063)
$A_3$ 's ranking: Top 5	0.765** (0.261)		
$A_3$ 's ranking: 6-15	0.548*** (0.153)		
$A_3$ 's ranking: 16-30	0.376** (0.141)		
$A_3$ 's ranking: 31-50	0.261* (0.127)		
$A_3$ 's ranking: 51-75	0.185 (0.114)		
$A_3$ 's ranking: 76-105	0.006 (0.116)		
$B_3$ 's ranking: Top 5	-0.765** (0.261)		
$B_3$ 's ranking: 6-15	-0.548*** (0.153)		
$B_3$ 's ranking: 16-30	-0.376** (0.141)		
$B_3$ 's ranking: 31-50	-0.261* (0.127)		
$B_3$ 's ranking: 51-75	-0.185 (0.114)		
$B_3$ 's ranking: 76-105	-0.006 (0.116)		
$A_3$ 's ranking vs $B_3$ 's ranking		X	
$\frac{Ranking_{A_3}}{Ranking_{B_3}} (< 1)$			-0.364** (0.134)
$A_3$ at home	-0.018 (0.127)	-0.039 (0.143)	-0.126 (0.114)
$B_3$ at home	0.018 (0.127)	0.039 (0.143)	-0.071 (0.123)
$A_3$ 's team ranking	-0.020 (0.011)	-0.019 (0.012)	-0.021* (0.011)
$B_3$ 's team ranking	0.020 (0.011)	0.019 (0.012)	0.015 (0.010)
Constant	0.513*** (0.142)	0.514** (0.161)	0.971*** (0.106)
<b>Controls for players' ability</b>			
Ranking category	YES	NO	NO
Interaction of rkg categories	NO	YES	NO
Ratio of rankings	NO	NO	YES
Observations	191	191	191
$R^2$	0.26	0.29	0.10

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 

## 2.5 Conclusion

Using team squash championships as a randomized natural experiment, we provide compelling empirical evidence against neutrality in multiple pairwise battles: in a best-of-three team contest, winning the first battle increases, *ceteris paribus*, the probability of winning the second battle. We show that this team dynamic is not driven by effects that would alter players' ability such as choking under pressure or psychological momentum. We further show that altruism and guilt aversion, two effects that affect players' prize spread, also fail to explain the absence of

neutrality.

The observed dynamic linkage is driven by another incentive effect, which we refer to as *individual contribution*: people derive utility from contributing to their team's success. To the best of our knowledge, this is the first paper providing empirical field evidence that individuals value contributing to their team's success. These findings cast new light on the way individuals behave in teams and open several potential avenues for future research.

While this effect should exist in other settings and in other kinds of contests outside the ones studied here, the subject merits further scrutiny. One important pending question is whether individual contribution depends on the observability of each teammate's performance. This has crucial implications with regard to management practices and contest design. If individual contribution only comes into play when performance is observable – as it is in the setting under scrutiny – organizations should design contests and team-based contracts in such a way that each teammate's performance is made public and the link between individual outcomes and team success is easily appreciable. For instance, the temporal structure of contests could play a major role in enabling organizations to reap the benefits of individual contribution, and sequential rounds could be much more efficient than simultaneous contests in inducing individuals to make a significant effort.

In any case, this effect is of prime importance to understanding team-based contests and contracts. Economic models of teams predict that individuals have an incentive to free ride, as they do not internalize the benefits accrued by other members of the team when making effort decisions.<sup>29</sup> Hence, the optimal level of effort exerted in individual contests should be higher than the effort observed in team-based incentive contests when the individual reward is based on team production. However, this theoretical result seems to contradict both experimental and behavioral literature on teams.<sup>30</sup> Individual contribution is a mechanism that mitigates free-riding behaviors. Each teammate values being responsible for collective success and therefore makes a significant effort, which could be higher than the effort that he would make in an individual contest. Individual contribution appears to be one of the reasons why so many firms use collective incentive and profit-sharing plans.

---

<sup>29</sup>See Lazear & Rosen (1981) for a seminal analysis of the relation between compensation and incentives in the presence of costly monitoring of worker's effort and output; and Prendergast (1999) for a survey of team production.

<sup>30</sup>See, for instance, Kandel & Lazear (1992). Chen & Lim (2013) show, using an experiment, that guilt is at play in teams and can explain why team-based incentives are more efficient than individual-based contracts.

## Appendices

### Appendix A - theoretical predictions

#### Neutrality

**T=3 (decisive battle)** Both teams won one individual battle. Players  $X_3$  and  $Y_3$  are now facing off in a decisive game. If  $X_3$  wins, which occurs with a probability  $\frac{x_3}{x_3+y_3}$ , he gets both the individual battle reward  $v$  and the collective reward  $V$ . If he loses, he gets a payoff 0. Whatever the result, he has to pay the cost of effort  $\frac{x_3}{\theta_{X_3}}$ .  $X_3$ 's maximization problem is therefore given by

$$\max_{x_3} \left( \frac{x_3}{x_3 + y_3} (v + V) - \frac{x_3}{\theta_{X_3}} \right).$$

Symmetrically, for player  $Y_3$ :

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} (v + V) - \frac{y_3}{\theta_{Y_3}} \right).$$

First-order conditions give the optimal levels of effort and  $p_3^*$ :

$$x_3^* = (v + V) \frac{\theta_{X_3}^2 \theta_{Y_3}}{(\theta_{X_3} + \theta_{Y_3})^2},$$

$$y_3^* = (v + V) \frac{\theta_{X_3} \theta_{Y_3}^2}{(\theta_{X_3} + \theta_{Y_3})^2},$$

$$p_3^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}.$$

This yields the following prediction:

**Prediction** The outcome of a decisive battle 3 depends solely on the relative ability of the players involved in battle 3.

**T=2** Team  $X$  won the first battle ( $X_1$  won against  $Y_1$ ).

$X_2$  chooses his level of effort  $x_2$  to maximize his utility. If he wins, which occurs with a probability  $\frac{x_2}{x_2+y_2}$ , he gets both the battle reward and the collective reward  $(v + V)$ . If he loses, which occurs with a probability  $\frac{y_2}{x_2+y_2}$ , he can still get the collective reward  $V$  if his teammate  $X_3$  wins the third battle (which occurs with a probability  $p_3^*$ ). Finally, whatever the outcome

of the battle, he has to pay the cost of his effort  $\frac{x_2}{\theta_{X_2}}$ .  $X_2$ 's maximization problem is therefore given by:

$$\max_{x_2} \left( \frac{x_2}{x_2 + y_2} (v + V) + \frac{y_2}{x_2 + y_2} p_3^* V - \frac{x_2}{\theta_{X_2}} \right).$$

$Y_2$  chooses his level of effort  $y_2$  to maximize his utility. If he wins, which occurs with a probability  $\frac{y_2}{x_2 + y_2}$ , he will get a payoff  $v + (1 - p_3^*)V$ . If he loses, the match ends and he gets a payoff 0. He has to pay the cost of effort  $\frac{y_2}{\theta_{Y_2}}$ , whatever the outcome of the battle. Note that in the neutrality model, the two players have the same prize spread (difference of utility between winning and losing):  $v + (1 - p_3^*)V$ .  $Y_2$ 's maximization problem is

$$\max_{y_2} \left( \frac{y_2}{x_2 + y_2} (v + (1 - p_3^*)V) - \frac{y_2}{\theta_{Y_2}} \right).$$

First-order conditions yield the optimal levels of effort and  $p_2^*$ :

$$x_2^* = (v + (1 - p_3^*)V) \frac{\theta_{X_2}^2 \theta_{Y_2}}{(\theta_{X_2} + \theta_{Y_2})^2};$$

$$y_2^* = (v + (1 - p_3^*)V) \frac{\theta_{X_2} \theta_{Y_2}^2}{(\theta_{X_2} + \theta_{Y_2})^2};$$

$$p_2^* = \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}.$$

This yields the two following predictions.

**Prediction 1** Winning battle 1 does not affect the probability of winning battle 2.

**Prediction 2** The outcome of battle 2 only depends on the two players involved in battle 2.

### Individual contribution

When a player wins and his team wins, he gets an additional payoff of individual contribution  $c$ .

	Player's team wins	Player's team loses
Player wins	$v + V + c$	$v$
Player loses	$V$	$0$

**T=3 (trivial battle)** Team  $X$  won battle 1 and battle 2, so battle 3 is trivial. If  $X_3$  wins, which occurs with a probability  $\frac{x_3}{x_3+y_3}$ , he gets a payoff  $v + V + c$  corresponding to the battle reward, the collective reward and the individual contribution effect. If he loses, he gets a payoff  $V$  (his team wins but he does not get the battle reward nor the individual contribution reward). Whatever the result, he has to pay the cost of effort  $\frac{x_3}{\theta_{X_3}}$ .

$$\max_{x_3} \left( \frac{x_3}{x_3 + y_3} (v + V + c) + \frac{y_3}{x_3 + y_3} V - \frac{x_3}{\theta_{X_3}} \right).$$

For player  $Y_3$ , the only stake is the battle reward  $v$ . As his team already lost, he is sure he will not get the collective reward, nor the individual contribution reward:

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} v - \frac{y_3}{\theta_{Y_3}} \right).$$

The first-order conditions give the optimal levels of effort and  $p_{3IC}^*$ :

$$x_{3IC}^* = \frac{(\theta_{X_3}(v+c))^2 \theta_{Y_3} v}{(\theta_{X_3}(v+c) + \theta_{Y_3} v)^2},$$

$$y_{3IC}^* = \frac{\theta_{X_3}(v+c)(\theta_{Y_3} v)^2}{(\theta_{X_3}(v+c) + \theta_{Y_3} v)^2},$$

$$p_{3IC}^* = \frac{\theta_{X_3}(v+c)}{\theta_{X_3}(v+c) + \theta_{Y_3} v}.$$

This yields the following prediction.

**Prediction** Winning battle 1 and battle 2 increases the probability of winning a trivial battle 3.

**T=3 (decisive battle)** Both teams won one individual battle. Players  $X_3$  and  $Y_3$  are now facing off in a decisive game. If  $X_3$  wins, which occurs with a probability  $\frac{x_3}{x_3+y_3}$ , his team wins and he gets a payoff  $v + V + c$ . If he loses, he gets a payoff 0. Whatever the result, he has to pay the cost of effort  $\frac{x_3}{\theta_{X_3}}$ .

$$\max_{x_3} \left( \frac{x_3}{x_3 + y_3} (v + V + c) - \frac{x_3}{\theta_{X_3}} \right).$$

Symmetrically for player  $Y_3$ :

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} (v + V + c) - \frac{y_3}{\theta_{Y_3}} \right).$$

The first-order conditions give the optimal levels of effort and  $p_{3IC}^*$ :

$$x_{3IC}^* = (v + V + c) \frac{\theta_{Y_3} \theta_{X_3}^2}{(\theta_{X_3} + \theta_{Y_3})^2};$$

$$y_{3IC}^* = (v + V + c) \frac{\theta_{X_3} \theta_{Y_3}^2}{(\theta_{X_3} + \theta_{Y_3})^2};$$

$$p_{3IC}^* = p_3^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}.$$

This yields the following prediction.

**Prediction** The outcome of a decisive battle 3 depends solely on the relative ability of the players involved in this battle.

**T=2** Team  $X$  won the first battle ( $X_1$  won against  $Y_1$ ). Contrary to T=3 (decisive battle), the two players do not face the same optimization problem.  $X_2$  chooses his level of effort  $x_2$  to maximize his utility. If he wins, he gets a payoff of  $v + V + c$ . If he loses, he gets a payoff of  $p_3^*V$  (he will get neither the private reward nor the “contribution reward” but he will get the collective reward if his teammate wins in T=3, which will occur with a probability  $p_3^*$ ). Therefore,  $X_2$ ’s maximization problem is given by:

$$\max_{x_2} \left( \frac{x_2}{x_2 + y_2} (v + V + c) + \frac{y_2}{x_2 + y_2} p_3^* V - \frac{x_2}{\theta_{X_2}} \right).$$

$Y_2$  chooses his level of effort  $y_2$  to maximize his utility. If he wins he gets a payoff of  $v + (1 - p_3^*)(V + c)$  because he will get the battle reward for sure and the collective reward and the individual contribution reward if his teammate wins, which will occur with a probability  $(1 - p_3^*)$ . If he loses he does not get any reward and ends up with a payoff 0.

$$\max_{y_2} \left( \frac{y_2}{x_2 + y_2} (v + (1 - p_3^*)(V + c)) - \frac{y_2}{\theta_{Y_2}} \right).$$

Thus,  $X_2$  has more incentive to win than  $Y_2$  because  $X_2$  is sure to get the “contribution reward” if he wins his battle while  $Y_2$  will get the “responsibility reward” if and only if his teammate also wins in T=3.

The first-order conditions yield the optimal levels of effort and  $p_{2IC}^*$ :

$$x_{2IC}^* = \frac{\theta_{X_2}^2 (v + (1 - p_3^*)V + c)^2 \theta_{Y_2} (v + (1 - p_3^*)(V + c))}{[\theta_{X_2} (v + (1 - p_3^*)V + c) + \theta_{Y_2} (v + (1 - p_3^*)(V + c))]^2},$$

$$y_{2IC}^* = \frac{\theta_{X_2} (v + (1 - p_3^*)V + c) \theta_{Y_2}^2 (v + (1 - p_3^*)(V + c))^2}{[\theta_{X_2} (v + (1 - p_3^*)V + c) + \theta_{Y_2} (v + (1 - p_3^*)(V + c))]^2},$$

$$p_{2IC}^* = \frac{\theta_{X_2} (v + (1 - p_3^*)V + c)}{\theta_{X_2} (v + (1 - p_3^*)V + c) + \theta_{Y_2} (v + (1 - p_3^*)(V + c))}.$$

As  $p_{2IC}^* > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$  and  $\frac{\partial p_{2IC}^*}{\partial p_3^*} > 0$ , this yields the two following predictions.

**Prediction 1** Winning battle 1 increases the probability of winning battle 2.

**Prediction 2** The probability of winning battle 2 increases with the probability of winning battle 3.

### Guilt aversion

When a player loses and his team loses, he gets a negative payoff payoff  $-s$ .

	Player's team wins	Player's team loses
Player wins	$v + V$	$v$
Player loses	$V$	$-s$

**T=3 (decisive battle)** Both teams won one individual battle. Players  $X_3$  and  $Y_3$  are now facing off in a decisive game. If  $X_3$  wins, which occurs with a probability  $\frac{x_3}{x_3 + y_3}$ , his team wins and he gets a payoff  $v + V$ . If he loses, he gets a payoff  $-s$  because he is “guilt-averse:” being partly responsible for the failure of his team is costly for him. Whatever the result, he has to pay the cost of effort  $\frac{x_3}{\theta_{X_3}}$  and faces the following maximization problem:

$$\max_{x_3} \left( \frac{x_3}{x_3 + y_3} (v + V) + \frac{y_3}{x_3 + y_3} (-s) - \frac{x_3}{\theta_{X_3}} \right).$$

Symmetrically for player  $Y_3$ :

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} (v + V) + \frac{x_3}{x_3 + y_3} (-s) - \frac{y_3}{\theta_{Y_3}} \right).$$

The optimal levels of effort and  $p_{3GA}^*$  are given by:



$$x_{3GA}^* = (v + V + s) \frac{\theta_{X_3}^2 \theta_{Y_3}}{(\theta_{X_3} + \theta_{Y_3})^2},$$

$$y_{3GA}^* = (v + V + s) \frac{\theta_{X_3} \theta_{Y_3}^2}{(\theta_{X_3} + \theta_{Y_3})^2},$$

$$p_{3GA}^* = p_3^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}.$$

This yields the following prediction.

**Prediction** The outcome of a decisive battle 3 depends solely on the relative ability of the players involved in this battle.

**T=2** Team  $X$  won the first battle ( $X_1$  won against  $Y_1$ ).  $X_2$  chooses his level of effort  $x_2$  to maximize his utility. If he wins, which occurs with a probability  $\frac{x_2}{x_2+y_2}$ , he gets a payoff  $v + V$ . If he loses, which occurs with a probability  $\frac{y_2}{x_2+y_2}$ , he gets a payoff  $V$  with a probability  $p_3^*$  and  $-s$  with a probability  $(1 - p_3^*)$ . Finally, whatever the outcome of the battle, he has to pay the cost of his effort  $\frac{x_2}{\theta_{X_2}}$ . His maximization problem is:

$$\max_{x_2} \left( \frac{x_2}{x_2 + y_2} (v + V) + \frac{y_2}{x_2 + y_2} (p_3^* V + (1 - p_3^*) (-s)) - \frac{x_2}{\theta_{X_2}} \right).$$

$Y_2$  chooses his level of effort  $y_2$  to maximize his utility. If he wins, which occurs with a probability  $\frac{y_2}{x_2+y_2}$ , he will get a payoff  $v + (1 - p_3^*)V$ . If  $Y_2$  loses, the match ends and he gets a payoff  $-s$ . He has to pay the cost of effort  $\frac{y_2}{\theta_{Y_2}}$ , whatever the outcome of the battle.

$$\max_{y_2} \left( \frac{y_2}{x_2 + y_2} (v + (1 - p_3^*)V) + \frac{x_2}{x_2 + y_2} (-s) - \frac{y_2}{\theta_{Y_2}} \right)$$

Thus,  $Y_2$  has more incentive to win than  $X_2$  because  $Y_2$  is sure to be “defeat-responsible” if he loses his battle while  $X_2$  will be “defeat-responsible” if and only if his teammate also loses in T=3.

Deriving the first-order conditions yield the optimal levels of effort and  $p_{2GA}^*$ :

$$x_{2GA}^* = \frac{\theta_{X_2}^2 (v + (1 - p_3^*)(V + s))^2 \theta_{Y_2} (v + (1 - p_3^*)V + s)}{[\theta_{X_2} (v + (1 - p_3^*)(V + s)) + \theta_{Y_2} (v + (1 - p_3^*)V + s)]^2},$$

$$y_{2GA}^* = \frac{\theta_{X_2} (v + (1 - p_3^*)(V + s)) \theta_{Y_2}^2 (v + (1 - p_3^*)V + s)^2}{[\theta_{X_2} (v + (1 - p_3^*)(V + s)) + \theta_{Y_2} (v + (1 - p_3^*)V + s)]^2},$$

$$p_{2GA}^* = \frac{\theta_{X_2}(v + (1 - p_3^*)(V + s))}{\theta_{X_2}(v + (1 - p_3^*)(V + s)) + \theta_{Y_2}(v + (1 - p_3^*)V + s)}.$$

As  $p_{2GA}^* < \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$  and  $\frac{\partial p_{2GA}^*}{\partial p_3^*} < 0$ , this yields the two following predictions:

**Prediction 1** Winning battle 1 decreases the probability of winning battle 2.

**Prediction 2** The probability of winning battle 2 decreases with the probability of winning battle 3.

### Disentangling individual contribution and altruism

We compare the predictions of individual contribution and altruism in battle 1 in the case where the favorites for battles 2 and 3 do not belong to the same team. Let  $X$  denote the team whose players are favorites in battle 2 and underdogs in battle 3. For simplicity, we furthermore assume that  $X_2$  will win with certainty and  $X_3$  will lose with certainty.<sup>31</sup>

**Individual contribution** Player  $X_1$  gets the battle reward  $v$ , the collective reward  $V$ , and the individual contribution reward  $c$  if he wins (as his teammate  $X_2$  will win battle 2 and end the contest) and he gets a payoff 0 if he loses. Player  $Y_1$  faces the same prize spread as he also gets  $V + v + c$  if he wins and 0 if he loses:

$$\Delta_{U_{X_1}} = \Delta_{U_{Y_1}} = v + V + c.$$

**Altruism** If players  $X_1$  and  $Y_1$  were altruistic, their incentives would no longer be symmetric. If  $X_1$  wins, he will get both the battle reward  $v$  and the collective reward  $V$  and he will prevent his teammate  $X_3$  from making a high effort in a decisive battle 3 (as the contest will be won after battle 2 thanks to the victory of  $X_2$ ). On the contrary, if  $X_1$  loses, he will get neither the battle reward nor the collective reward and he will force his teammate  $X_3$  to play, which induces a negative payoff  $-\alpha C(X_3)$  where  $C(X_3)$  is the cost of effort of  $X_3$  in a decisive battle 3 and  $\alpha$  reflects the degree to which  $X_1$  internalizes this cost ( $0 < \alpha < 1$ ). Hence the prize spread of  $X_1$  will be  $v + V + \alpha C(X_3)$ . His opponent  $Y_1$  faces a different problem. If he wins, he gets both  $v$  and  $V$  as his team will win the contest but he forces his teammate  $Y_3$  to play a decisive battle

---

<sup>31</sup>Note that the logic would be the same with a more general framework where  $X_2$  is “as much of a favorite as  $X_3$  is an underdog”.

3, which is partly internalized by him ( $-\alpha C(Y_3)$ ). If  $Y_1$  loses, he gets neither  $v$  nor  $V$  but he prevents  $Y_3$  from playing. Hence his prize spread is  $v + V - \alpha C(Y_3)$ .

$$\Delta_{U_{X_1}} = v + V + \alpha C(X_3) > \Delta_{U_{Y_1}} = v + V - \alpha C(Y_3)$$

**Different predictions** Thus, individual contribution predicts that  $X_1$  and  $Y_1$  have the same prize spread while altruism predicts that  $X_1$  has a higher prize spread than  $Y_1$ . Since  $p_1^* = \frac{\theta_{X_1} \Delta_{U_{X_1}}}{\theta_{X_1} \Delta_{U_{X_1}} + \theta_{Y_1} \Delta_{U_{Y_1}}}$ , individual contribution predicts that being the favorite in battle 2 is equivalent to being the favorite in battle 3 whereas altruism predicts that being the favorite in battle 2 is preferable to being the favorite in battle 3. This finding is the basis for our empirical test in section 2.4.2.

### Choking under pressure

**T=3 (trivial battle)** Team  $X$  already won battles 1 and 2, so battle 3 becomes stakeless (except for the battle reward) and neither of the two players faces pressure. Hence their cost of effort are not affected.

Player  $X_3$  is sure to get the collective reward  $V$  whatever the outcome of battle 3 but he will get the battle reward  $v$  only if he wins battle 3.

$$\max_{x_3} \left( V + \frac{x_3}{x_3 + y_3} v - \frac{x_3}{\theta_{X_3}} \right)$$

Player  $Y_3$  will not get the collective reward but he can get the battle reward  $v$  if he wins.

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} v - \frac{y_3}{\theta_{Y_3}} \right)$$

First-order conditions give the optimal levels of effort and  $p_{3CUPtrivial}^*$ :

$$x_{3CUPtrivial}^* = v \frac{\theta_{X_3}^2 \theta_{Y_3}}{(\theta_{X_3} + \theta_{Y_3})^2},$$

$$y_{3CUPtrivial}^* = v \frac{\theta_{X_3} \theta_{Y_3}^2}{(\theta_{X_3} + \theta_{Y_3})^2},$$

$$p_{3CUPtrivial}^* = p_3^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}.$$

This yields the following prediction.

**Prediction** Winning battle 1 and battle 2 does not affect the probability of winning a trivial battle 3.

**T=3 (decisive battle)** Both team won one individual battle. Players  $X_3$  and  $Y_3$  are now opposed in a decisive game. Since battle 3 is pivotal, players  $X_3$  and  $Y_3$  might both choke under pressure as losing the battle implies the defeat of their team. This is conceptually equivalent to multiplying their ability by  $\eta$  with  $0 < \eta < 1$ .

$$\max_{x_3} \left( \frac{x_3}{x_3 + y_3} (v + V) - \frac{x_3}{\theta_{X_3} \eta} \right)$$

Symmetrically for player  $Y_3$ :

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} (v + V) - \frac{y_3}{\theta_{Y_3} \eta} \right)$$

Deriving the FOCs yield the optimal levels of effort and  $p_3^*$ :

$$x_{3CUPdecisive}^* = (v + V) \eta \frac{\theta_{X_3}^2 \theta_{Y_3}}{(\theta_{X_3} + \theta_{Y_3})^2}$$

$$y_{3CUPdecisive}^* = (v + V) \eta \frac{\theta_{X_3} \theta_{Y_3}^2}{(\theta_{X_3} + \theta_{Y_3})^2}$$

$$p_{3CUPdecisive}^* = p_3^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}$$

This yields the following prediction.

**Prediction** The outcome of a decisive battle 3 depends solely on the relative ability of the players involved in this battle.

**T=2** Team  $X$  won the first battle ( $X_1$  won against  $Y_1$ ). There is an asymmetry between the two players. Player  $X_2$  has no reason to choke under pressure because his team will remain in the contest if he loses battle 2. On the contrary, player  $Y_2$  knows that losing battle 2 implies that his team loses the contest. Thus, he may choke under pressure, which can be modelled by multiplying his ability by a factor  $\eta$  ( $0 < \eta < 1$ ).

$$\max_{x_2} \left( \frac{x_2}{x_2 + y_2} (v + V) + \frac{y_2}{x_2 + y_2} p_3^* V - \frac{x_2}{\theta_{X_2}} \right).$$

$$\max_{y_2} \left( \frac{y_2}{x_2 + y_2} (v + (1 - p_3^*)V) - \frac{y_2}{\theta_{Y_2}\eta} \right).$$

First-order conditions yield the optimal levels of effort and  $p_2^*$ :

$$x_{2CUP}^* = (v + (1 - p_3^*)V) \frac{\theta_{X_2}^2 \theta_{Y_2} \eta}{(\theta_{X_2} + \theta_{Y_2} \eta)^2};$$

$$y_{2CUP}^* = (v + (1 - p_3^*)V) \frac{\theta_{X_2} (\theta_{Y_2} \eta)^2}{(\theta_{X_2} + \theta_{Y_2} \eta)^2};$$

$$p_{2CUP}^* = \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2} \eta}$$

As  $p_{2CUP}^* > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$ , this yields the following prediction.

**Prediction** Winning battle 1 increases the probability of winning battle 2.

### Psychological momentum

**T=3 (decisive battle)** Team  $X$  won the first battle and lost the second battle ( $X_1$  won against  $Y_1$  and  $X_2$  lost against  $Y_2$ ).  $Y_3$  has psychological momentum because his teammate won the previous battle. This is conceptually equivalent to multiplying his ability by a factor  $\psi$  (with  $\psi > 1$ ). The maximization problem is:

$$\max_{x_3} \left( \frac{x_3}{x_3 + y_3} (v + V) - \frac{x_3}{\theta_{X_3}} \right),$$

$$\max_{y_3} \left( \frac{y_3}{x_3 + y_3} (v + V) - \frac{y_3}{\theta_{Y_3} \psi} \right).$$

Optimal levels of effort and  $p_{3PM}^*$  are therefore given by:

$$x_{3PM}^* = (v + V) \frac{\theta_{X_3}^2 (\theta_{Y_3} \psi)}{(\theta_{X_3} + \theta_{Y_3} \psi)^2},$$

$$y_{3PM}^* = (v + V) \frac{\theta_{X_3} (\theta_{Y_3} \psi)^2}{(\theta_{X_3} + \theta_{Y_3} \psi)^2},$$

$$p_{3PM}^* = \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3} \psi}.$$

As  $p_{3PM}^* < \frac{\theta_{X_3}}{\theta_{X_3} + \theta_{Y_3}}$ , this yields the following prediction.

**Prediction** In a decisive battle 3, the player in the team that won battle 2 is more likely to win than the player in the team that won battle 1.

**T=2** Team  $X$  won the first battle ( $X_1$  won against  $Y_1$ ).  $X_2$  has psychological momentum because his teammate won the previous battle. This is conceptually equivalent to multiplying his ability by a factor  $\psi$  (with  $\psi > 1$ ). We therefore have the following maximization problems:

$$\max_{x_2} \left( \frac{x_2}{x_2 + y_2} (v + V) + \frac{y_2}{x_2 + y_2} p_{3PM}^* V - \frac{x_2}{\theta_{X_2} \psi} \right),$$

$$\max_{y_2} \left( \frac{y_2}{x_2 + y_2} (v + (1 - p_{3PM}^*) V) - \frac{y_2}{\theta_{Y_2}} \right).$$

Deriving the first order conditions yield the optimal levels of effort and  $p_{2PM}^*$ :

$$x_{2PM}^* = (v + (1 - p_{3PM}^*) V) \frac{(\theta_{X_2} \psi)^2 \theta_{Y_2}}{(\theta_{X_2} \psi + \theta_{Y_2})^2};$$

$$y_{2PM}^* = (v + (1 - p_{3PM}^*) V) \frac{(\theta_{X_2} \psi) \theta_{Y_2}^2}{(\theta_{X_2} \psi + \theta_{Y_2})^2};$$

$$p_{2PM}^* = \frac{\theta_{X_2} \psi}{\theta_{X_2} \psi + \theta_{Y_2}}.$$

As  $p_{2PM}^* > \frac{\theta_{X_2}}{\theta_{X_2} + \theta_{Y_2}}$ , this yields the following prediction.

**Prediction** Winning battle 1 increases the probability of winning battle 2.

## Appendix B - probit estimation

Table 2.12: Evidence against neutrality (probit estimation - average marginal effects)

	(1)	(2)	(3)	(4)	(5)	(6)
$A_1$ won battle 1	0.105*** (0.025)	0.073** (0.027)	0.117*** (0.027)	0.081** (0.029)	0.083*** (0.024)	0.063* (0.026)
<b>Controls for players' ability</b>						
Ranking category	YES	YES	NO	NO	NO	NO
Interaction of rkg categories	NO	NO	YES	YES	NO	NO
Ratio of rankings	NO	NO	NO	NO	YES	YES
<b>Other controls</b>						
Playing home + teams' rankings	NO	YES	NO	YES	NO	YES
Observations	934	896	934	896	934	896

Standard errors in parentheses: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Reading note (column 1): the average marginal effect of winning battle 1 on the probability of winning battle 2 is 10.5 pp.

Note: the number of observations decreases from 934 in specifications (1), (3), (5) to 896 in specifications (2), (4), (6)

because there are 38 contests for which teams' rankings are missing.

Table 2.13: Evidence against neutrality (probit estimation - full table).

	(1)	(2)	(3)	(4)	(5)	(6)
<b>A<sub>1</sub> won battle 1</b>	0.440*** (0.106)	0.308** (0.115)	0.457*** (0.108)	0.317** (0.118)	0.363*** (0.108)	0.278* (0.116)
A <sub>2</sub> 's ranking: Top 5	3.071*** (0.278)	2.672*** (0.315)				
A <sub>2</sub> 's ranking: 6-15	2.483*** (0.222)	2.159*** (0.251)				
A <sub>2</sub> 's ranking: 16-30	1.911*** (0.194)	1.647*** (0.217)				
A <sub>2</sub> 's ranking: 31-50	1.172*** (0.176)	0.998*** (0.191)				
A <sub>2</sub> 's ranking: 51-75	0.735*** (0.188)	0.587** (0.199)				
A <sub>2</sub> 's ranking: 76-105	0.378* (0.188)	0.292 (0.198)				
B <sub>2</sub> 's ranking: Top 5	-3.071*** (0.278)	-2.672*** (0.315)				
B <sub>2</sub> 's ranking: 6-15	-2.483*** (0.222)	-2.159*** (0.251)				
B <sub>2</sub> 's ranking: 16-30	-1.911*** (0.194)	-1.647*** (0.217)				
B <sub>2</sub> 's ranking: 31-50	-1.172*** (0.176)	-0.998*** (0.191)				
B <sub>2</sub> 's ranking: 51-75	-0.735*** (0.188)	-0.587** (0.199)				
B <sub>2</sub> 's ranking: 76-105	-0.378* (0.188)	-0.292 (0.198)				
A <sub>2</sub> 's ranking vs B <sub>2</sub> 's ranking			X	X		
$\frac{Ranking_{A_2}}{Ranking_{B_2}} (< 1)$					-2.120*** (0.202)	-1.998*** (0.229)
A <sub>2</sub> at home		0.034 (0.210)		0.040 (0.215)		0.410 (0.238)
B <sub>2</sub> at home		-0.034 (0.210)		-0.040 (0.215)		0.182 (0.212)
A <sub>2</sub> 's team ranking		-0.043** (0.016)		-0.044** (0.016)		-0.020 (0.017)
B <sub>2</sub> 's team ranking		0.043** (0.016)		0.044** (0.016)		0.033* (0.015)
Constant	-0.220 (0.140)	-0.154 (0.228)	-0.228 (0.163)	-0.158 (0.259)	1.647*** (0.147)	1.428*** (0.176)
<b>Controls for players' ability</b>						
Ranking category	YES	YES	NO	NO	NO	NO
Interaction of rkg categories	NO	NO	YES	YES	NO	NO
Ratio of rankings	NO	NO	NO	NO	YES	YES
Observations	934	896	934	896	934	896

Standard errors in parentheses: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Note: the number of observations decreases from 934 in specifications (1), (3), (5) to 896 in specifications (2), (4),(6) because there are 38 contests for which teams' rankings are missing.





## Chapter 3

# Estimating individual productivity in football

JEAN-BAPTISTE VILAIN & RODRIGO LOPEZ KOLKOVSKY

## **Abstract**

In most production settings, the collective output is observed while the individual productivity of team members is difficult to assess. We develop an estimation procedure based on fixed-effects and a Ridge penalization in order to address this issue. We focus on football because it represents an ideal laboratory: the rules of the game are clearly defined and data are publicly available. Once our productivity metrics obtained, we confront it to players' market value in order to identify undervalued and overvalued players on the market. We find significant statistical evidence that the proportion of black players is greater in the group of undervalued players than in the group of overvalued players.

**JEL Classification** C81, Z20, Z22.

**Keywords** Team Economics, Individual Productivity, Sports Analytics, Ridge, Discrimination.

### 3.1 Introduction

Assessing individual productivity in teamwork is very complex because only the collective output can be directly observed. The marginal productivity of team members is difficult to measure, especially when teammates' efforts interact in a subtle manner (Alchian & Demsetz (1972)).

Team sports are no exception: sportsmen are a special type of workers that are associated together on the field to produce a collective output. This collective output is easily observed through the games' outcomes (goals, points, etc) but it is difficult to know precisely to what extent players contribute to the success of their team. Of course team sports offer more information on individual performance than most other team settings because games can be watched. Therefore players' actions are observable and it is possible to have a vague idea about players' individual performance. However, subjective assessment on players' performance tends to be systematically biased by the *scoring illusion*: the impact of offensive and spectacular tasks tend to be overstated while defensive and unspectacular tasks like marking, that are equally important for the team, tend to be largely understated (Kuper (2014)).

Estimating the individual level of a sportsman in a team is a challenging exercise because the level of the opponents and the level of the teammates need to be taken into account. As it is more difficult for a player to make a difference against a high performing team, an excellent performance against a strong team must be weighted more than an excellent performance against a weak team. Similarly, a player is likely to better play if he is surrounded by teammates who place him in good playing conditions. Therefore, the individual productivity of a player must be netted out of his teammates' productivity.

Two different approaches have been considered in the literature to estimate individual productivity in team sports. The first approach is based on *individual "box-score" statistics* such as dribbles, passes or shots. It is a two-stage procedure. First, actions that make the team successful are identified by regressing the team output<sup>1</sup> on the different box-score statistics. Each action gets a coefficient according to its contribution to the the success of the team. Then, individual productivity is obtained by multiplying these coefficients by the number of actions of each player. This method has the advantage of giving a lot of information on the aspects of the game where players contribute the most, so that it is possible to understand why a player is productive. However, the box-score approach might fail to take into account all the

---

<sup>1</sup>The winning percentage is an example of a team output that might be considered in the box-score approach.

dimensions of players' contribution to the team's success. Some key individual statistics might not be identified and might not appear in the box-score. For example, some defensive statistics like the marking ability are difficult to record. Furthermore, personality traits like charisma or leadership, that are strongly related to productivity, cannot be measured.

The second approach used in the literature does not rely on any in-game statistics other than players' participation. It corresponds to a *fixed-effect* estimation: a player's productivity is estimated according to his participation to different games. The better the results obtained by a team when a player is on the field, the higher his productivity is.<sup>2</sup> Thus the fixed effect approach captures all the dimensions of players' productivity. It also has the advantage of being very low data-demanding as only the final scores and the line-ups for every game are needed in the estimation. The main limit of this method is that, contrary to the box-score approach, it is completely agnostic about how a player contributes to the success of his team. Hence it does not provide any insight on what makes a player productive.

Among team sports, football<sup>3</sup> is particularly difficult to analyze. Contrary to baseball which can be decomposed into discrete sequences where a batter is opposed to a pitcher, football is continuous due to the nearly uninterrupted flow of the ball during the game. Moreover, relevant performance statistics are not clearly identified in football, contrary to other sports such as basketball where statistics on points, rebounds, steals, assists, blocks, turnovers or fouls are very informative on players' performance. Finally, scoring events are very rare in football (on average there are less than 3 goals scored during a game), so considering simple statistics such as the number of goals or assists is not very representative of the level of a player. These aspects make football analytics particularly challenging. It also explains why measures of individual productivity developed in other team sports cannot be directly transposed to football (see section 3.2).

The inherent complexity of football makes this setting similar to most work environments. Indeed, in most firms: (i) it is often difficult to decompose a work flow in different discrete sequences; (ii) relevant performance variables at the individual level may be hard to identify; (iii) variability in global performance of the team may be limited. Hence, football exhibits the same characteristics as the majority of team settings and appears as an ideal laboratory. Our

---

<sup>2</sup>That is why we term it a *fixed-effect approach* even if it is not necessary a fixed-effect approach in the usual econometric sense.

<sup>3</sup>From now on, when we use the word football, we mean European football (soccer) and not American football.

first contribution is to use football data in order to develop an estimation procedure of individual productivity in teams. Our methodology combines fixed-effects with a Ridge penalization. Fixed-effects allow us to consider individual productivity in all its dimensions while the Ridge penalization prevents overfitting as it imposes a constraint on the sum of squared estimated coefficients. It also allows us to overcome the colinearity issue due to the fact that players often play with the same teammates.

We obtain results that are consistent with specialists' opinion. Lionel Messi and Cristiano Ronaldo, who are largely recognized as the two best players over the period under consideration, top our ranking. All other players in our top 30 are world class players. This suggests that our estimation procedure is very well-suited to the analysis of individual productivity in teams and could be applied to a wide scope of settings outside football. The only requirements are the observability of the collective output and frequent shifts between teammates/workers (see section 3.6.2).

Our second contribution deals with the functioning of the labor market. In most empirical studies, labor market discrimination is measured by estimating wage functions for various groups of workers. Discrimination is then operationally defined as differences in predicted wages for the different groups when the prediction holds constant various productivity determinants of wages (see Cain (1986) for a survey). However, interpreting differences in predicted wages by groups as discrimination may be problematic for two reasons. First, as pointed out by Neal & Johnson (1996), resulting estimates of residual wage gaps may be biased if some controls for worker productivity are endogenous, i.e. are themselves affected by market discrimination. This is the case of controls such as occupation, postsecondary schooling, part-time work, marital status, geographical location, and actual labor market experience, which are all subject to worker choice and could be contaminated by labor market discrimination.<sup>4</sup> Second, data on worker productivity are often limited, so controls for worker productivity may be very imprecise measures of worker skills. Hence, it might be difficult to distinguish labor market outcomes arising from discrimination against a group from those produced by intergroup differences in unobserved productivity.

The impact of these problems should be limited in our setting. First, it seems reasonable to

---

<sup>4</sup>See Corcoran & Duncan (1979); Reimers (1983); Smith & Welch (1986); O'Neill (1990); Blau & Beller (1992); Oaxaca & Ransom (1994).

assume that our measure of individual productivity is not itself affected by discrimination. As our estimation procedure is based on players' fixed effects and only relies on data on final scores and line-ups, it seems unlikely that the bias introduced by the Ridge penalization is correlated to players' race. Second, labor market outcomes of football players should be largely determined by their on-field productivity, so intergroup differences in unobservables such as education should not alter much the analysis. Thus, we can test for racial discrimination by simply comparing labor market outcomes of white and black players at equivalent level of estimated productivity. Szymanski (2000) finds evidence of lower wages for black players in English football leagues. His identification strategy is based on the fact that if football were a competitive non-discriminatory market for the services of players, clubs' performance should only depend on their wage bill. Yet, he shows that clubs fielding an above-average proportion of black players systematically outperform clubs with a below-average proportion of black players, after controlling for the wage bill. Confronting our productivity metrics to players' market value is another approach to test whether there is discrimination on the football market. It is complementary to Szymanski's study for two reasons. First, we consider a different labor market outcome: we focus on players' market values on the transfer market instead of their wages. Second, and more importantly, our test for discrimination is realized at the player's level thanks to our individual productivity metrics. By comparing the proportion of black players among undervalued and overvalued players, we provide evidence of racial discrimination on the football market. The share of black players is about three times higher in the group of undervalued players than in the group of overvalued players.

The remainder of this paper is organized as follows. Section 3.2 presents in more details the sports analytics literature on the box-score and fixed-effect approaches. Section 3.3 presents our data and our estimation procedure. Our productivity rankings and our market test for racial discrimination are exposed in section 3.4. In section 3.5, we use in-sample and out-of-sample simulations to test the robustness of our measure of individual productivity. Section 3.6 presents two potential extensions of our estimation procedure. Section 3.7 concludes.

## 3.2 Related literature

In the literature on sports analytics, two different approaches are used to evaluate individual productivity in team sports: the *box-score* approach and the *fixed-effect* approach.

### 3.2.1 The “box-score” approach

Berri (1999) and Berri & Schmidt (2010) are very illustrative of the box-score approach. First, they identify the actions that make the success of a basketball team (points scored, rebounds etc). Then, they measure individual productivity by looking at how many of these actions the players make. Each action is weighted accordingly to its importance for winning. The measure they get is extremely simple<sup>5</sup> but it does not take into account opponents’ level. Franck & Nüesch (2010) adapt the Berri approach to European football. They start by estimating the impact of different actions on the winning percentage of teams. The productivity of a player is then defined as the sum of his individual actions multiplied by their impact on the teams’ winning percentage.

Another example of the box-score approach is the use of the DEA (data envelopment analysis). Depending on some inputs and some outputs, an efficiency frontier is obtained. Players who lie on the efficiency frontier are the most efficient, i.e. those who produce the most outputs relatively to the inputs they use. Tiedemann, Francksen & Latacz-Lohmann (2011) use a DEA to compare football players. They use playing time as an input and goals scored, assists, the percentage of successful passes and the percentage of successful tackles as outputs. Hence, they implicitly assume that the productivity of a player only depends on these four variables.

### 3.2.2 The fixed-effect approach

The main idea behind the fixed-effect approach is that box-score statistics cannot take into consideration all the dimensions of a player’s contribution to his team’s success. A broader measure, that also reflects the unobservable part of productivity, is needed.

In the fixed-effect approach, a player’s productivity is estimated according to his participation to different games. Müller, Upmann & Prinz (2013) use a fixed-effect approach based on network analysis. They model a team as a network and players as nodes. Edges between two players depend on their performance when they are together on the field. For example, if player A and player B get on average 2 points when they play together while player A gets on average 2.5 points (unconditionally on player B’s participation) and player B gets on average 1.5 points (unconditionally on player A’s participation), the edge from player A to player B will be 0.8 (2/2.5) and the edge from player B to player A will be 1.25 (2.5/2). The stronger the edges that

---

<sup>5</sup> $WinScore = Points + Rebound + Steals + 0.5Assists + 0.5Blocks - Turnovers - FieldGoalsAttempted - 0.5Fouls - 0.5FreeThrowsattempted$



arrive to a player, the more his teammates outperform when he is on the field. The authors define the adjacency matrix by computing these numbers for every pair of players and assess players' individual productivity by their eigenvector centrality in the network. Their approach seems appealing for it measures players' contributions to results, whether tangible or intangible. However, as the network is defined as the team, this method does not allow to compare the productivity of players from different teams. Altman (2013) introduces the Shapley value to football to assess how pivotal a player is in the goal difference. Like eigenvector centrality, Shapley values are agnostic about how the actions of players translate into performance. The main limit of this method is that results may be sensitive to the low variability in teams' squads and the low number of goals scored and conceded.

A simpler example of a fixed-effect approach is the *plus-minus*. This measure simply compares the goal difference when a player is on the field to the goal difference when he is not on the field. The *plus-minus* has been progressively improved in order to be adjusted for the level of teammates and the level of opponents. Rosenbaum (2004) proposes a version of an adjusted plus-minus in basketball. He defines an observation as a sequence of a game in which there is no shift. The dependent variable is the goal difference of the home team. The explanatory variables are dummy variables for every player. A dummy is equal to 0 if a player is not on the field, to 1 if he is on the field and plays for the home team and to -1 if he is on the field and plays for the away team. Ilardi & Barzilai (2008) suggest another version of adjusted plus-minus in basketball that allows to differentiate between offensive and defensive productivity of the players. Any sequence of a game opposing team A and team B during which there is no shift is decomposed into two observations: the attacking ability of team A's players against the defensive ability of team B's players and the attacking ability of team B's players against the defensive ability of team A's players. The dependent variable is the number of points scored by the attacking players and the explanatory variables are dummies  $X_i$  for attacking players and  $D_j$  for defending players. Each player has thus both an offensive and a defensive productivity.

In basketball, the variability of points scored during a sequence without shift is very high and the substitutions are frequent, which gives room for estimation of the adjusted plus-minus. However, in other sports where scoring events are much rarer like ice-hockey and football, variability of the data does not seem sufficient for the standard adjusted plus-minus to work well. The obtained coefficients have a very large variance and there is a problem of overfitting. One way of overtaking this issue is to use penalization methods as in Gramacy, Jensen & Taddy

(2013) and Thomas, Ventura, Jensen & Ma (2013) who estimate the productivity of ice-hockey players. Penalization methods are indeed very effective in settings where the number of variables is very high and where there is strong colinearity (see Section 3.3.3)

### 3.3 Data and estimation strategy

#### 3.3.1 Data

##### Data used in the estimation of players' individual productivity

We collect, clean, merge and format data from four different websites<sup>6</sup> in order to build a comprehensive dataset on the final scores and lineups for all the games of the seasons 2007/2008 to 2014/2015 in which two teams from the Big 5<sup>7</sup> are opposed. All the games of the five national leagues, part of the national cups games and part of the European cups games are thus taken into account.<sup>8</sup> This makes a total of 16006 games over the eight seasons. For every game, we have data on the home team, the final score, players' positions and players' playing time. We do not have information about the time at which shifts occur during the game, so we consider entire games rather than sequences of games without shift.

Since the sample consists in all the games in which two clubs from the Big 5 are opposed, all the players that played in a Big 5 club between 2007/2008 and 2014/2015 are considered in our analysis.<sup>9</sup> Doing the estimation on such a broad sample rather than on one league only is a way of making the data less unbalanced and to get more variability as players are often transferred from one league to another. In total, 3866 players are present in our sample.

We modify the structure of the dataset. We decompose every game into two observations. If team  $I$  plays against team  $J$ , we decompose it into one observation with the offensive players of team  $I$  against the defensive players of team  $J$  and one observation with the offensive players

---

<sup>6</sup>Data on final scores can be found on the following websites: <http://www.football-data.co.uk/> (national leagues), <http://fr.uefa.com/> (European games) and <http://fr.soccerway.com/> (national cups games). Data on players' participation are publicly available on [www.transfermarkt.co.uk/](http://www.transfermarkt.co.uk/)

<sup>7</sup>English Premier League, Spanish Liga, Italian Serie A, German Bundesliga and French Ligue 1

<sup>8</sup>The precise list of competitions considered is the following : Premier League, Liga, Serie A, Bundesliga, Ligue 1, Europa League, Champions League, Community Shield, FA Cup, League Cup, Copa del Rey, Coppa Italia, Supercoppa Italiana, DFB-Pokal DFL-Supercup, Coupe de France, Coupe de la Ligue, Trophée des champions.

<sup>9</sup>We estimate the productivity of players whose playing time is at least 15 games over the entire period. When their playing time is lower than 15 games, we group them into team fixed-effects. For example, in a 2014/2015 Ligue 1 game, if Guingamp has a lineup with two forwards who played less than 15 games over the period, we put them together in  $\theta_{Guingamp2014/2015}$ . The associated playing time to  $\theta_{Guingamp2014/2015}$  for this game is 180 (twice 90 minutes). By doing so, we implicitly assume that the level of different players from the same team who have a very low playing time is the same. This assumption is necessary to get some room for estimation.

of team  $J$  against the defensive players of team  $I$ . This transformation allows us to disentangle players' offensive and defensive productivity, as in the adjusted plus-minus approach of Ilardi & Barzilai (2008). We end up with a total of 32012 observations.

Besides, we collect data on the pre-game odds (win/draw/loss) used by four betting operators for all the games of the five domestic leagues that took place between 2007/2008 and 2014/2015.<sup>10</sup> From the odds, we can compute the underlying probabilities used by the operators in their forecasting models and compare them to our predictions in order to assess the robustness of our estimation (see section 3.5).

### Data used in the market test for discrimination

Two labor market outcomes can be considered for a football player: his wage and his market value, *i.e.* the amount of money a club needs to pay to transfer the player. Data on individual wages are confidential and not publicly available, so we cannot confront our measure of productivity to players' wages. Data on individual market values of players can be obtained on the reputable football website *Transfermarkt*. The "true" market value of players who are not transferred or who are loaned out are unobserved but the Transfermarkt values appear as a very good proxy for them.<sup>11</sup> We collect annual market values for all players in our sample.

In order to implement our market test for racial discrimination, we also need to collect data on players' skin color. Following Deschamps & De Sousa (2014), this information is coded from an examination of players' photographs into categories of either black or not black. This method might sound arbitrary but it is actually a good way to model the potential for discrimination because discriminators prejudge an individual based on appearances. We record information on skin color for the top 150 forwards in our sample. We restrict our analysis to this subsample of very-high achieving players because data on players' race must be recorded manually and we want to ensure that players' pictures are easily accessible on different websites.

### 3.3.2 The structural equation

Consider a game opposing team  $I$  and team  $J$ . The number of goals scored by team  $I$  to team  $J$  will depend positively on team  $I$ 's players' offensive productivity, and negatively on team  $J$

---

<sup>10</sup>Those are publicly available on the website <http://www.football-data.co.uk/>.

<sup>11</sup>Transfermarkt is a German website specialized in assessing the market value of players. Most of the time, when a player is transferred, the actual fees are very close to the Transfermarkt value. See <http://www.transfermarkt.co.uk/>

players' defensive productivity, weighted by their respective playing times:

$$y_{IJ}^* = \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i - \sum_{j \in J_I} t_{j_{JI}} \times \delta_j + \beta_{Home} \times Home_I + \epsilon_{IJ} \quad (3.1)$$

Where

$y_{IJ}^*$  is the latent variable associated to  $y_{IJ}$ , the number of goals scored by team  $I$  to team  $J$

$i \in I_J$  if player  $i$  from team  $I$  is on the field during the game against team  $J$

$j \in J_I$  if player  $j$  from team  $J$  is on the field during the game against team  $I$

$t_{i_{IJ}}$  is the playing time of player  $i$  during the game between team  $I$  and team  $J$

$t_{j_{JI}}$  is the playing time of player  $j$  during the game between team  $I$  and team  $J$

$\theta_i$  is the offensive productivity of player  $i$

$\delta_j$  is the defensive productivity of player  $j$

$Home_I$  is a dummy variable equal to one if team  $I$  plays at home

$\beta_{Home}$  reflects the playing home advantage

$\epsilon_{IJ}$  is a random shock that follows a standard normal distribution

$\theta_i$ s,  $\delta_j$ s and  $\beta_{Home}$  are the coefficients to be estimated.

According to our model :

- The number of goals scored by team  $I$  to team  $J$  increases with the offensive productivity of team  $I$ 's players. We observe different combinations of players associated to different numbers of goals scored. If team  $I$  scores a lot every time player  $A$  is on the field (controlling for his teammates and opponents), player  $A$  will have a high  $\theta_A$ . Therefore variations of the lineups from one game to the next is the source of identification for the coefficients  $\theta$ s. Transfers, injuries and turnover in general are sources of variations that facilitate the identification of the coefficients  $\theta$ s.
- Football is a setting where there is opposition between two teams. Therefore, the number of goals scored by team  $I$  to team  $J$  also depends on the defensive productivity of team  $J$ 's players, so as to weight performance according to the difficulty of scoring. In that sense, our model is very similar to Rasch (1993) used in psychometrics where the probability of a correct answer depends positively on the respondent's ability and negatively on the difficulty of the question asked.<sup>12</sup> In our setting, variations of the lineups are again the

---

<sup>12</sup>In the Rasch model, the probability of a correct response is given by  $P(Y_{ik} = 1|\theta_i, b_k) = (1 + \exp(b_k - \theta_i))^{-1}$

source of identification for the coefficients  $\delta$ s. If a team concedes a really low number of goals every time player  $B$  is on the field (controlling for his teammates and opponents), then player  $B$  will have a high  $\delta_B$ .

- The number of goals scored by team  $I$  to team  $J$  also depends on the playing home advantage (that is used here as a control variable) and on a random shock.

Given its form, our model has the advantage of netting out individual productivity from the level of the opposition and the level of teammates. Thus, good performance against very good players will induce larger player productivity coefficients than the same performance against weaker players. In the same spirit, good performance with weak teammates will induce larger player productivity coefficients than the same performance with strong teammates.

The coefficients  $\theta$ s and  $\delta$ s can be interpreted as player fixed effects that reflect both the pure technical individual ability of players and their ability to make their teammates play better. For example, a forward may have a high  $\theta$  either because he scores a lot of goals or because he enables the other forward in his team to score a lot. This externality on teammates would be difficult to capture if we considered only individual statistics of players such as goals and shots. In this case, only the scorer would be rewarded.

Note that we could use other metrics such as ball possession, the number of shots or shots' accuracy instead of the number of goals scored in our structural equation. This would give us more information on aspects of the game where each player contributes the most. Unfortunately, data on these metrics at the team level are not publicly available for most competitions in our sample so we could not test this approach.

### 3.3.3 Estimation procedure

#### Generalities

The estimation of the structural equation (3.1) can be done through a variety of techniques which include least squares, logit regression, probit regression and poisson regression. Considering that the variable of interest, the number of goals scored per match  $y_{IJ}$ , is a discrete nonnegative variable, the prime candidates are the likelihood methods and the poisson regression. The poisson regression, however, involves the assumption that times of arrival of goals are independent of each other, which seems unlikely in football.

---

for individual  $i$  with ability level  $\theta_i$  and item difficulty parameter  $b_k$ .

We will therefore estimate the parameters of interest via maximum likelihood, and more particularly in the ordered probit regression. The main principle of this technique is that the outcome is contingent of the latent continuous variable  $y_{IJ}^*$  falling between different thresholds  $\{\alpha_i\}_{i=1}^k$ , as in (3.2):

$$y_{IJ} = \begin{cases} 0 & \text{if } y_{IJ}^* \leq \alpha_1 \\ k & \text{if } \alpha_k < y_{IJ}^* \leq \alpha_{k+1} \quad \text{and} \quad k \in 1, \dots, K-1 \\ K & \text{if } y_{IJ}^* > \alpha_K \end{cases} \quad (3.2)$$

The estimation procedure relies on the assumption that the residual terms of (3.1) follow a standard normal distribution, thus enabling a maximum likelihood procedure.

### Main challenges

One of the main difficulties of estimation is the risk of overfitting, which consists in the existence of biases on the coefficient estimates caused by idiosyncratic patterns in the data which are strongly influenced by chance rather than the systematic relationship of interest (Bilger, Manning et al. (2011)). This problem is exacerbated when the model is particularly complex and the number of observations is limited, leading to good in-sample predictions but poor performance on external data sets. In our particular framework, overfitting poses an important risk for two reasons:

- The number of coefficients to estimate is very high as we have one or two variables per player. This makes a total of more than 4400 coefficients to estimate, so we are clearly in a *large p* setting.
- There is a strong colinearity issue, as players often are on the field with the same teammates. Thus, it is hard to isolate their individual contribution from the contribution of their teammates.

### Limitations imposed to the coefficients' dynamics

Having a very realistic model is of course desirable but it is in general not compatible with a feasible estimation. Considering both the computational difficulties of the estimation algorithm as well as the data size, we need to introduce limitations to the coefficients' dynamics. Three assumptions are necessary to reduce the number of coefficients to be estimated and to make the

estimation tractable:

Assumption 1: players' productivities stay constant over the observed timespan. Indeed, allowing the player fixed effects to vary by season would dramatically increase the number of variables to be estimated and make the estimation procedure cumbersome. Relaxations of this assumption with the use of player specific ageing curves are discussed in section 3.6.

Besides, as productivities are fixed, we do not take into account the fact that players may exert different levels of effort depending on the circumstances of the game (e.g. the fact that players might make less effort when the outcome of the game is stakeless for the championship). These kind of effects are too fine-grained to be captured here but they are investigated in a different setting in the second chapter of this dissertation.

Assumption 2: players' productivities do not depend on specific teammates. Interactions terms (e.g.  $\theta_1 * \theta_2$ ) are not included in the model because the number of coefficients to be estimated is already extremely large. Adding interactions makes the estimation untractable, even with shrinkage. Effectively, in a framework which includes thousands of players, the inclusion of first order interactions would potentially imply the inclusion of millions of columns to the data matrix which renders the estimation computationally impossible.

Assumption 3: specialization of tasks by position. Forwards only have an offensive productivity, midfielders have both an offensive and a defensive productivity and defenders only have a defensive productivity (table 3.1).<sup>13</sup> A player's position is considered fixed over time and is defined as his most frequent position over the estimation time span.

	<b>Attacking productivity <math>\theta</math></b>	<b>Defensive productivity <math>\delta</math></b>
<b>Forward</b>	✓	x
<b>Midfielder</b>	✓	✓
<b>Defender</b>	x	✓

Table 3.1: Offensive and defensive productivity by position

### Estimation of player fixed-effects with shrinkage

Even under these simplifying assumptions, traditional estimation methods lead to coefficients

---

<sup>13</sup>We do not take into account goalkeepers in the estimation because the turnover is way lower for this position, which worsens the colinearity issue.

that seem to explain the data very well but are not robust to out of sample tests, casting doubt on the validity of the estimates in terms of statistical significance as well as in terms of prediction properties. One of the most popular techniques to deal with the issue of high colinearity is the Ridge Regression, introduced by Hoerl & Kennard (1970), which consists in introducing an L2-norm penalty of the coefficients to the least squares penalisation criteria. This technique can effectively reduce the mean square error of the predicted response at the cost of biasing the estimates towards zero. In our case, the Ridge objective function then becomes:

$$L(\theta, \delta|y, t_{ij}) = \left( \sum_{IJ} \sum_{k=0}^K \mathbb{1}_{y_{IJ}=k} \times Pr(y_{IJ} = k) \right) - \left( \lambda \sum_z (\theta_z^2 + \delta_z^2) \right) \quad (3.3)$$

The function to be maximized  $L$  has two components :

- The first component  $\sum_{IJ} \sum_{k=0}^K \mathbb{1}_{y_{IJ}=k} \times Pr(y_{IJ} = k)$  is a simple log-likelihood function (ordered probit). Maximizing this part of  $L$  is equivalent to maximizing the goodness of fit. The precise expression for  $Pr(y_{IJ} = k)$  is given in Appendix A.
- The second component  $\lambda \sum_z (\theta_z^2 + \delta_z^2)$  is the penalization/shrinkage part of the function  $L$ . Note that the coefficient  $\beta_{Home}$  is not penalized in our estimation. This is justified by the fact that it is clearly established that playing home has a big impact on teams' success.

Thus, the optimization procedures trades off between the goodness of fit and the penalization component. The higher  $\lambda$ , the more the penalization component will be important and the more coefficients will be biased towards zero. Two extreme cases are  $\lambda = 0$  and  $\lambda = \infty$ . In the case  $\lambda = 0$ , there is no penalization and the optimization will only take into account goodness of fit (simple ordered probit). In the case  $\lambda = \infty$ , all the coefficients will be equal to zero and the goodness of fit will not matter at all. To understand the mechanism induced by the L2-norm penalty more generally, it is useful to consider simpler frameworks. In the case of the linear model, it can be shown that for sufficiently large  $\lambda$ , the optimisation problem is not singular regardless of the explanatory variables being perfectly collinear. Furthermore, by using a Singular Value Decomposition, it can be shown that the coefficients of variables which are strongly associated to principal components with larger eigenvalues will be shrunked in a weaker way and that perfectly colinear variables will be assigned the same coefficient (see Friedman, Hastie & Tibshirani (2001)). This implies that coefficients of explanatory variables with a large predictive power will be more impervious to shrinkage than variables that explain a small fraction



of the variance and that grouped variables will receive a common coefficient. Therefore, players who play often together will receive similar coefficients, which contrasts with other Shrinkage estimators such as the Lasso.

An important question when implementing a composite objective function is the number of critical points in its domain, and whether there exists a unique global maximum. Effectively, the optimisation of objective functions that possess several local maxima may prove difficult as the use of traditional methods is sensible to the choice of initial parameters. To circumvent this issue, it is often appealing to use concave objective functions, as they assure that the algorithm will reach a unique global maximum and avoid staying in local maxima. In particular, concave objective functions are less dependent to the choice of starting values for the optimisation routine. In our case, the objective function is composed of an ordered probit likelihood function minus a L2 norm of the parameters. The ordered probit likelihood function is concave as long as the participation matrix does not have colinear columns, and the minus L2 norm is a globally concave function, thus assuring that the optimisation problem has a unique solution and that estimates are robust to the choice of initial search parameters.

The penalisation parameter  $\lambda$  plays a crucial role in both the interpretation and prediction accuracy of the estimates. Large penalisation weights will lead to coefficients heavily shrunked towards zero and most of the predicted values will be close to the average value of the dependent variable. However, it can be shown that there exists some  $\lambda > 0$  which will always reduce the RMSE with respect to a model where  $\lambda = 0$ .  $\lambda$  is often chosen through a model selection technique, such as cross-validation. Explicitly, the  $K$ -fold cross validation procedure consists in randomly partitioning the data into  $K$  subsets of comparable size; one of the subsamples is taken as the test set, while the rest is taken as the training set. The model's parameters are estimated on the training set and their prediction accuracy is tested on the test set. This procedure is repeated for each of the  $K$  subsets and the residuals are combined to produce a summary statistic such as the Mean Squared Error. We perform a 10-fold cross-validation and we choose the penalisation parameter so as to minimize the mean squared difference between the observed number of goals and the expected number of goals derived from the estimated coefficients.

The penalisation parameter  $\lambda$  can be interpreted by the use of Bayesian statistics. Indeed, including the penalisation term is equivalent to assuming that coefficients  $(\theta, \delta)$  have a prior

distribution which is normal with mean zero and variance  $\lambda$ . Large values of lambda therefore imply the assumption that there are strong reasons to believe that there is little differentiation between players' skills.

## Implementation

We coded the estimation procedure on R.<sup>14</sup>

## 3.4 Results

### 3.4.1 Interpreting the coefficients

Our estimation relies on an ordered probit model. This implies that the estimated coefficients  $\hat{\theta}$  and  $\hat{\delta}$  are difficult to interpret. In order to obtain a productivity measure that is easier to interpret, we convert the estimated coefficient of a player into his marginal goal contribution when his teammates and opponents are average players of the Big 5.<sup>15</sup>

We start by calculating the expected number of goals scored by a team of average players that is opposed to another team of average players. Assuming that both teams play with 4 defenders, 3 midfielders and 3 forwards (4-3-3), we get an expected number of goals of 1.32.<sup>16</sup>

Then we look at the marginal contribution of players, i.e. the variations of the expected number of goals scored and conceded when an average player is replaced by the player of interest. For concreteness, we consider the case of Gonzalo Higuain. To obtain his marginal goal contribution, we consider again a game opposing average players and we substitute one of them by Higuain. This yields an expected number of goals of 1.52, so the offensive marginal contribution of Higuain is 0.2 goals (1.52 - 1.32). The principle is the same for defensive marginal effects. Substituting an average player by John Terry in a game opposing only average players yield an expected number of goals conceded of 1.16 so the defensive marginal contribution of Terry is 0.16 goals (1.32-1.16).

We do this operation for every player and we get a measure expressed in marginal goal contribution per game that is easily interpretable. Since midfielders have both an offensive and

---

<sup>14</sup>Penalization methods are not widely used with ordered models and there is no pre-coded package on statistical softwares for this kind of estimation. Hence, we coded the estimation procedure by ourselves.

<sup>15</sup>Note that the introduction of the L2 penalty in the optimization implies that the marginal effects are biased. For the ease of interpretation, we do not take into consideration this issue and we assume that the bias is the same for every player.

<sup>16</sup>The expected number of goals is given by  $E(y_{IJ}) = \sum_{k=0,1,2,\dots,K} k \times Pr(y_{IJ} = k)$

a defensive contribution, their aggregate marginal effect is defined as the sum of their offensive and defensive marginal contributions.

### 3.4.2 The most productive players

Table 3.2 reports the top 25 players according to our estimation. Their marginal contribution reflects their individual productivity in Big 5 clubs from 2007/2008 to 2014/2015.

Table 3.2: Overall ranking of players: top 25 (all positions)

Ranking	Player	Total	Offensive	Defensive
1	Lionel Messi	.351	.351	
2	Cristiano Ronaldo	.333	.333	
3	Cesc Fabregas	.302	.243	.059
4	Frank Lampard	.274	.205	.069
5	Yaya Toure	.255	.138	.117
6	Karim Benzema	.253	.253	
7	Bastian Schweinsteiger	.246	.171	.075
8	Xavi	.241	.175	.066
9	Arturo Vidal	.238	.128	.11
10	Franck Ribery	.237	.237	
11	Gonzalo Higuain	.235	.235	
12	Mesut Ozil	.228	.228	
13	Andres Iniesta	.224	.13	.094
14	Luka Modric	.218	.125	.093
15	Arjen Robben	.217	.217	
16	Kwadwo Asamoah	.194	.103	.091
17	Wayne Rooney	.186	.186	
18	Marco Verratti	.183	.118	.065
19	Thomas Muller	.182	.182	
20	Zlatan Ibrahimovic	.181	.181	
21	Eden Hazard	.181	.181	
22	Luis Suarez	.179	.179	
23	Javier Pastore	.178	.178	
24	Ivan Rakitic	.176	.087	.089
25	Pedro	.176	.176	

Seasons 2007/2008 to 2014/2015. Lambda = 200. Only midfielders have both an offensive and a defensive contribution.

This ranking seems quite consistent with the general perception of players' productivity. Lionel Messi and Cristiano Ronaldo, who have both won 5 times the Ballon d'Or<sup>17</sup> top the

<sup>17</sup>The Ballon d'Or is the most famous annual MVP award in football.

ranking. Cesc Fabregas, Franck Lampard, Yaya Toure, Karim Benzema, Bastian Schweinsteiger, Xavi, Arturo Vidal and Franck Ribéry complete the top 10 of our ranking. These eight players were all world class players over the period. Other players who were generally also considered as overachievers such as Andres Iniesta, Arjen Robben Wayne Rooney, Thomas Muller, Eden Hazard, Zlatan Ibrahimovic or Luis Suarez rank in the top 25. Our estimation relies only on a fixed-effects model where players' productivity is assessed according to their team's performance when they are on the field, controlling for their teammates and their opponents. The model is completely agnostic about what makes a player productive. Thus, it is very satisfying to get a ranking with our statistical method that is very close to the "common knowledge" based on people perceptions.

No defender appears in the overall top 25. The first defender, John Terry, ranks only 30th in the overall ranking (see the top 100 ranking in Appendix B). This finding seems consistent with the observed transfers fees by position. Defenders are usually much cheaper than midfielders and forwards on the market, which might be explained by a higher substitutability.

Tables 3.4, 3.6 and 3.8 report the top 20 players by position. Karim Benzema follows Lionel Messi and Cristiano Ronaldo in our ranking of forwards. The French striker was named UNFP Ligue 1 Player of the Year in 2007/2008 thanks to his performance in Lyon. After another season in Lyon, Benzema was transferred to Real Madrid where he became an essential player. He was named French Player of the Year three times in 2011, 2012 and 2014. Franck Ribery ranks 4th. The winger has won six Bundesliga titles, five DFB-Pokal, one UEFA Champions League and one FIFA Club World Cup with Bayern Munich. Thanks to his exceptional performance in 2012/2013, he was nominated alongside Lionel Messi and Cristiano Ronaldo on the three-man shortlist for the 2013 FIFA Ballon d'Or. Gonzalo Higuain completes our top 5. He scored 107 goals in 190 league appearances with Real Madrid before joining Napoli where he became the club's star.

Table 3.4: Forwards' ranking: top 20

Ranking	Player	ME_total	ME_offensive	ME_defensive
1	Lionel Messi	.351	.351	
2	Cristiano Ronaldo	.333	.333	
3	Karim Benzema	.253	.253	
4	Franck Ribery	.237	.237	
5	Gonzalo Higuain	.235	.235	
6	Mesut Ozil	.228	.228	
7	Arjen Robben	.217	.217	
8	Wayne Rooney	.186	.186	
9	Thomas Muller	.182	.182	
10	Zlatan Ibrahimovic	.181	.181	
11	Eden Hazard	.181	.181	
12	Luis Suarez	.179	.179	
13	Javier Pastore	.178	.178	
14	Pedro	.176	.176	
15	Robert Lewandowski	.174	.174	
16	Edin Dzeko	.169	.169	
17	Alexis Sanchez	.164	.164	
18	Sergio Aguero	.159	.159	
19	Mario Gotze	.156	.156	
20	Kaka	.148	.148	

Seasons 2007/2008 to 2014/2015.

Cesc Fabregas tops our ranking of midfielders. This is not surprising since the Spanish player managed to become a starter in Arsenal when he was only 16 years old. Then, he broke other club's records, earning a reputation as one of the best players in his position. Frank Lampard, who is second in our ranking, is the all-time leading goalscorer for Chelsea, where he played for 13 years. He is considered by a number of journalists and football experts to be one of the greatest midfielders of his generation. Yaya Toure, 3rd in our ranking, was voted African Footballer of the Year for 2011, 2012, 2013 and 2014. He is followed by Bastian Schweinsteiger, who spent 13 seasons at Bayern Munich, the best football club in Germany. His honours at the club include eight Bundesliga titles, seven DFB-Pokal titles, a UEFA Champions League title, a FIFA Club World Cup title and a UEFA Super Cup title. Xavi, the last player in our top 5, has been awarded the IFFHS World's Best Playmaker award for four years: 2008, 2009, 2010, 2011. He has also been included in the UEFA Team of the Year for five years (2008, 2009, 2010,

2011, 2012).

Table 3.6: Midfielders' ranking: top 20

Ranking	Player	ME_total	ME_offensive	ME_defensive
1	Cesc Fabregas	.302	.243	.059
2	Frank Lampard	.274	.205	.069
3	Yaya Toure	.255	.138	.117
4	Bastian Schweinsteiger	.246	.171	.075
5	Xavi	.241	.175	.066
6	Arturo Vidal	.238	.128	.11
7	Andres Iniesta	.224	.13	.094
8	Luka Modric	.218	.125	.093
9	Kwadwo Asamoah	.194	.103	.091
10	Marco Verratti	.183	.118	.065
11	Ivan Rakitic	.176	.087	.089
12	Paul Scholes	.171	.125	.045
13	Dani Parejo	.165	.133	.033
14	Tiago	.163	.091	.071
15	Darren Fletcher	.162	.108	.054
16	Asier Illarramendi	.155	.112	.043
17	Sulley Muntari	.155	.066	.089
18	Nuri Sahin	.149	.097	.053
19	Yohan Cabaye	.143	.079	.063
20	Thiago	.141	.117	.023

Seasons 2007/2008 to 2014/2015.

According to our productivity measure, John Terry is the best defender over the period we consider. This result seems consistent with the great performance of this player. He has been Chelsea's most successful captain, having led them to five Premier League titles, four FA Cups, three League Cups, one UEFA Europa League and one UEFA Champions League. He was also named UEFA Club Defender of the Year in 2008 and 2009. Rafinha was an important player in Schalke 04, Genoa and Bayern Munich during our estimation time span. Even though he performed very well in these three clubs, it is surprising to find him in second position in our ranking, as he is usually not considered as one of the best defender in the world. Nemanja Vidic completes our top 3. The Serbian player collected a host of honours in his Manchester United career, including five Premier League titles, one UEFA Champions League and three League Cup medals. At the individual level, he helped United to a record-breaking run of 14

consecutive clean sheets in 2008/2009 and was awarded the Premier League Player of the Season award. Giorgio Chiellini won six consecutive Serie A titles with Juventus Turin, as well as three consecutive Coppa Italia titles, and three Supercoppa Italiana titles. He is considered as one of the best defender in the world as he was ranked as the seventh-best footballer playing in Europe by Bloomberg in 2013. The fifth player in our ranking, Medhi Benatia, performed very well in Udinese, AS Roma and Bayern Munich over the time span.

Table 3.8: Defenders' ranking: top 20

Ranking	Player	ME_total	ME_offensive	ME_defensive
1	John Terry	.165		.165
2	Rafinha	.165		.165
3	Nemanja Vidic	.153		.153
4	Giorgio Chiellini	.149		.149
5	Medhi Benatia	.14		.14
6	Juanfran	.139		.139
7	Jeremy Toulalan	.139		.139
8	Dante	.132		.132
9	Xabi Alonso	.121		.121
10	Gerard Pique	.121		.121
11	Jerome Boateng	.12		.12
12	Maxwell	.118		.118
13	Alvaro Arbeloa	.118		.118
14	Jamie Carragher	.117		.117
15	Thiago Silva	.114		.114
16	Philipp Lahm	.111		.111
17	Javier Mascherano	.111		.111
18	Miranda	.11		.11
19	Loïc Perrin	.108		.108
20	Vincent Kompany	.106		.106

Seasons 2007/2008 to 2014/2015.

### 3.4.3 Comparing our measure of productivity to the market value of players

A simple way to test our measure of players' individual productivity is to confront it to players' market value over our estimation time span.<sup>18</sup> Since individual productivity is the key variable

<sup>18</sup>As a player's market value is different from one season to the next, we consider the average market value over the seasons where the player plays in a Big 5 club, *i.e.* the seasons considered for the player in our estimation.

that impacts players' value,<sup>19</sup> we should observe a strong correlation between our coefficients and players' market values.

The relationship between the two measures is clearly positive, which shows that our estimation of individual productivity is globally aligned with the market (figure 3.1). Moreover, the convexity of the trend indicates that marginal productivity is increasingly expensive, a finding consistent with the economics of superstars (Rosen 1981): Lionel Messi and Cristiano Ronaldo are only slightly more productive than other top forwards but this slight difference has a huge impact on their market value.

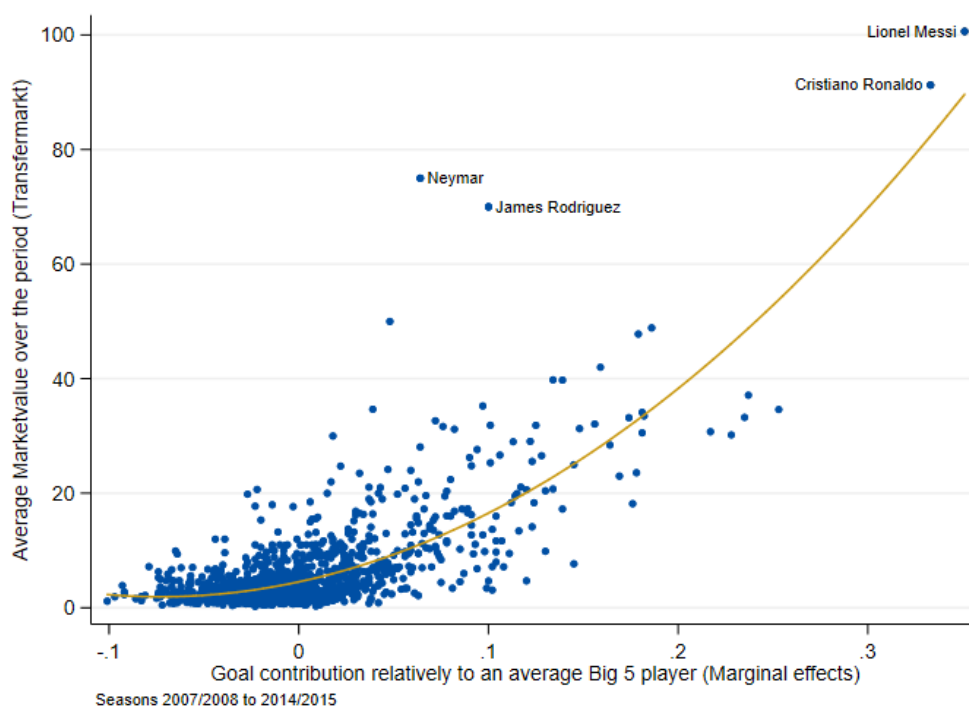
Neymar and James Rodriguez are two outliers in our analysis. Their average market values seem disproportionate compared to their productivity. This can be explained by the fact that we only take into account the "on-field" productivity of players and not the externalities they generate "off-field" such as sponsoring contracts or merchandising products. Since Neymar and James are world stars, the amount of money they generate off-field may compensate the gap between their productivity on-field and their market value.

---

<sup>19</sup>Players' market values also depend on players' age (at equivalent performance, young players are more expensive than experienced players because they are expected to be enrolled in the team on a longer period), contract duration (the longer a player is engaged in a club, the more expensive he is for another club) and clubs involved in the transfer (all clubs do not have the same financial resources and the same power of negotiation).



Figure 3.1: Productivity vs. Market value (analysis on forwards)



### 3.4.4 A market test for discrimination

In most empirical settings, it is very difficult to identify discrimination from observational data. Regressing the labor market outcome under consideration (usually wage or employment status) on productivity related variables available in the data and on individual characteristics subject to potential discrimination (usually race or gender) may not allow to identify precisely the degree of discrimination. As unobservables can be correlated to the characteristic subject to potential discrimination, the estimated labor market differential may not reflect discrimination. This problem is well illustrated in the review on racial discrimination of Lang & Lehmann (2012). They show that the magnitude and the significance of the wage differential between blacks and whites in the US are very sensitive to the productivity related variables included in the estimation. In a first study Neal & Johnson (1996) find that after controlling for age and performance on the Armed Forces Qualifying Test (AFQT), the black–white wage differential among young men is modest (about 7%) and statistically insignificant. However, other studies obtain different results by controlling for additional predictors of wages. Rodgers & Spriggs (1996) and Carneiro, Heckman & Masterov (2005) find that including years of schooling at the time the respondents

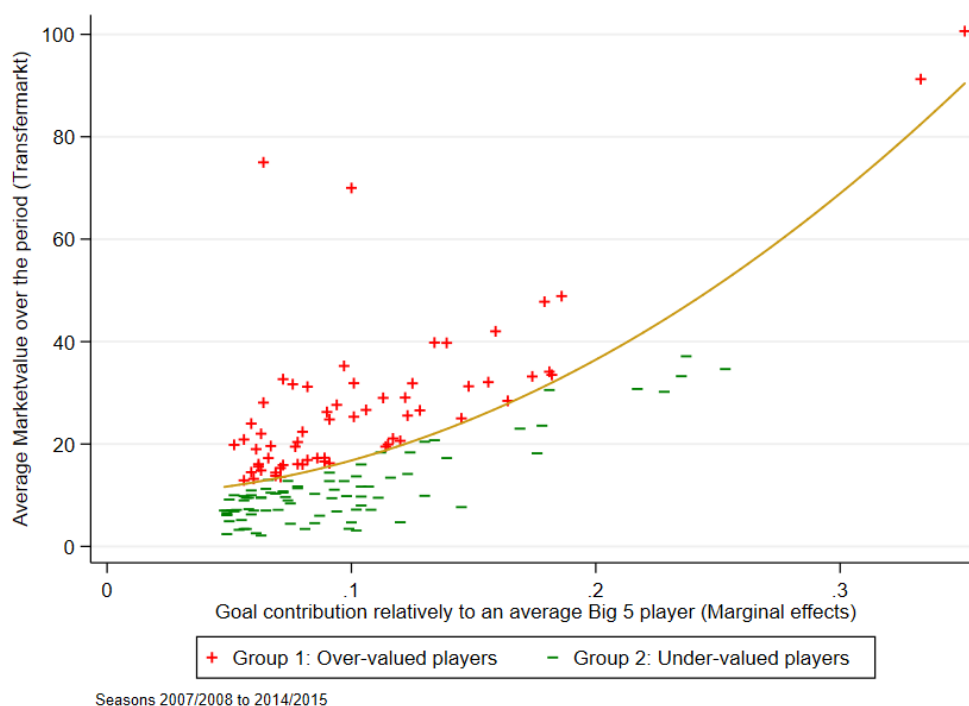
took the AFQT leads to a substantial wage differential. When controlling for final education attainment, Lang & Manove (2011) reach a similar conclusion because conditional on AFQT, blacks get more education than whites. Their result can be challenged by the fact that blacks attend lower quality schools on average. Black, Kolesnikova, Sanders & Taylor (2013) find that controlling for location also increases the estimated wage gap between blacks and whites. These different examples show that results are strongly affected by the productivity variables that are accounted for. As it is usually impossible to get data sets with all the factors impacting productivity, the estimated wage differential does not necessarily reflect discrimination.

This problem should be limited in our setting for it seems reasonable to consider that football players' labor market outcomes are largely determined by their productivity on-field. Unobserved factors such as education, that could vary by groups/race, should have a very limited impact on players' labor market outcomes. Therefore, the magnitude of the bias should be very small and we can test for discrimination by simply confronting our measure of on-field productivity to players' labor market outcomes. Figure 3.2 plots the market value against our productivity metrics for the top 150 forwards in our sample.<sup>20</sup> We can identify two groups of players: those who are undervalued and those who are overvalued on the market. Our criterion to determine to which group a player belongs is his position relative to the quadratic fit of the market value on productivity. Players above the fit are considered overvalued, while players below the fit are considered undervalued.

---

<sup>20</sup>Note that we restrict our market test for discrimination on this subsample of 150 players because data on players' race must be recorded manually and we want to ensure that pictures of the players are accessible on different websites.

Figure 3.2: Identifying overvalued and undervalued players (analysis on top 150 forwards)



Our market test for discrimination consists in assessing whether the share of black players is greater in the group of undervalued players than in the group of overvalued players. We perform a standard statistical test on the equality of proportions of black players in the two groups (see table 3.10). Black players account for 24.4% of players in the group of undervalued players whereas they represent only 9.4% of overvalued players. The difference of proportion is statistically significant at the 5% level ( $p$  value = 0.018). This is clear evidence that black players tend to be more often undervalued on the market than other players.

Table 3.10: Market test for discrimination: comparison of the share of black players among undervalued and overvalued players

Group	N	Proportion of black players	p-value
Under-valued players	86	0.244	
Over-valued players	64	0.094	
Difference between the two groups		0.150*	0.018

\* Statistically significant at the 5% level.

Our results are in line with Szymanski (2000), who finds evidence of lower wages for black players in English leagues. Here, we focus on a different labor market outcome since we consider players' market value instead of their wages. Market values do not directly involve players' remuneration, as they correspond to the transfer fees that should be paid by a club to another club to engage one of its player. Therefore, the fact that black players are undervalued on the transfer market does not directly imply that their own wages are lower and that they are directly discriminated. However, from a theoretical viewpoint, it seems difficult to imagine a mechanism according to which black players are undervalued on the "transfer market" without being also undervalued on the "salary market". The fact that clubs' total spending in transfer fees are strongly correlated to their total payrolls confirms empirically that such an asymmetric mechanism is unlikely. Thus, our analysis provides indirect evidence of discrimination against black players.

## 3.5 Robustness

### 3.5.1 In-sample predictions with simulations of seasons

One way of testing the robustness of our measure of productivity is to simulate games according to the scoring probabilities of teams derived from the estimated coefficients. If the measure is reliable, the predicted scoring probabilities should be reliable too and the average results obtained in simulations should be close to the observed results.

#### Principle of the simulations

The probabilities of scoring can be easily derived from the coefficients  $\theta_s$ ,  $\delta_s$ ,  $\beta_{Home}$  and cut-off values estimated in the ordered probit model (see Appendix A). From the probabilities of scoring, we can derive the probabilities of win, draw and loss. We use the probabilities of win, draw and loss rather than the probabilities of scoring because the aim of a team is to maximize points rather than goals. Therefore, it makes more sense to make predictions on points.

Consider a game between team  $I$  and team  $J$ . Let  $y_{IJ}$  be the number of goals scored by team  $I$  to team  $J$  :

$$Pr(I \text{ wins}) = Pr(J \text{ loses}) = \sum_{k=1,2,\dots,K} Pr(y_{IJ} = k) \times Pr(y_{JI} < k)$$

$$Pr(I \text{ draws}) = Pr(J \text{ draws}) = \sum_{k=0,1,2,\dots,K} Pr(y_{IJ} = k) \times Pr(y_{JI} = k)$$

$$Pr(I \text{ loses}) = Pr(J \text{ wins}) = \sum_{k=0,1,2,\dots,K} Pr(y_{IJ} = k) \times Pr(y_{JI} > k)$$

Then, we can simulate games according to these probabilities. In a simulation, team  $I$  will win with a probability  $Pr(I \text{ wins})$ , will draw with a probability  $Pr(I \text{ draws})$  and will lose with a probability  $Pr(I \text{ loses})$ . We aggregate the results of the simulation at the season level for each of the five domestic leagues. For example in simulation 1, if a team wins 20 games, draws 8 times and loses 10 times in its league over the period, our prediction will be 68 points. Then we can rank teams in their championship according to the simulation. Of course, every simulation yields a different outcome. Therefore, we do 300 different simulations and we derive an average ranking of teams in their league. For example, if a team ranks  $3^{rd}$  in 100 simulations,  $4^{th}$  in 100 simulations and  $5^{th}$  in 100 simulations, our average predicted ranking will be 4th. We can assess the goodness of fit of our estimation procedure by comparing our average predicted ranking to the observed ranking of teams.

### Results of the simulations : goodness of fit

The fit between the observed ranking and our average predicted ranking is really good as most of the dots lie next to the 45 degree line (figures 3.3 and 3.4). This is strong evidence in favor of our measure :

- It shows that the aggregation of players' individual productivity reflects pretty well the collective productivity of teams.
- Given the structural equation (3.1), a statistical error in the estimation of the productivity of a player will induce a statistical error of the opposite sign in the productivity coefficient of his teammates. For example, if two players always play together, an upward bias in the coefficient of one of them can be compensated with a downward bias in the coefficient of the other one, thus leading to the same fit when aggregating by teams.<sup>21</sup> Hence the fact that

---

<sup>21</sup>To see this, take the example of players  $A$  and  $B$  who always play together: the coefficients  $(\theta_A, \theta_B)$  acquire the same value of the likelihood function as  $(\theta_A + \mu, \theta_B - \mu)$ , meaning that we cannot identify these coefficients in the non-shrunked model.

the aggregation of individual productivities reflects well the collective productivity does not necessarily mean that the coefficients of individual productivity are well estimated.

This problem is mitigated when players move frequently across teams: an error in the estimated coefficient of any player would induce an error in the coefficients of his teammates in every club where he has played, which would result in poor prediction regarding collective productivity. Considering that players are often transferred (on average, they play in 2 different teams and in 1.2 different leagues over the eight seasons span) and that our predictions of collective performance fit the data well, our estimates are unlikely to suffer from this identification problem.

Figure 3.3: Goodness of fit : seasons 2007/2008 to 2010/2011

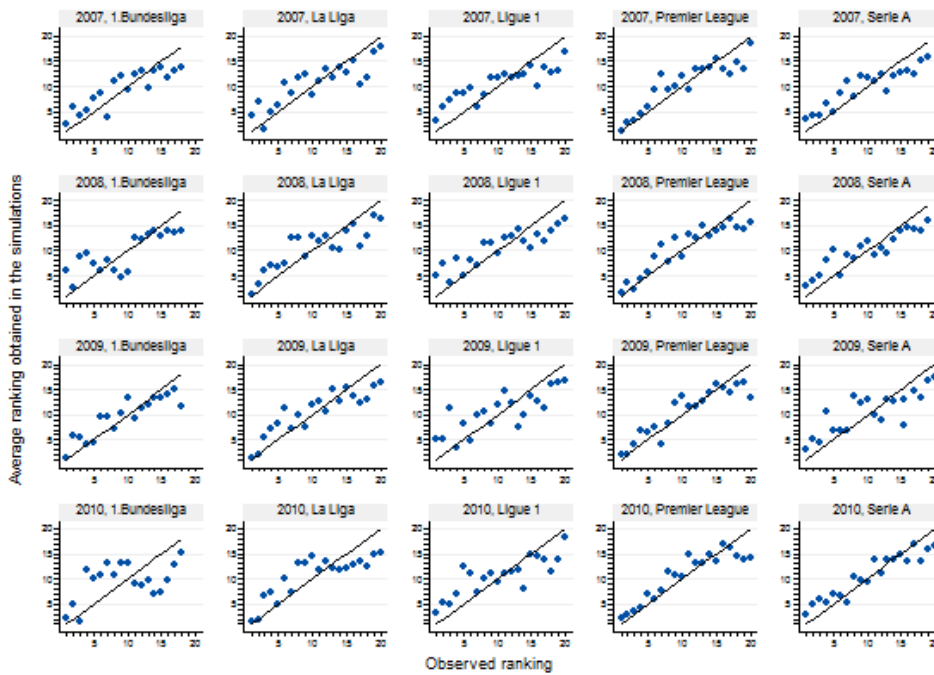
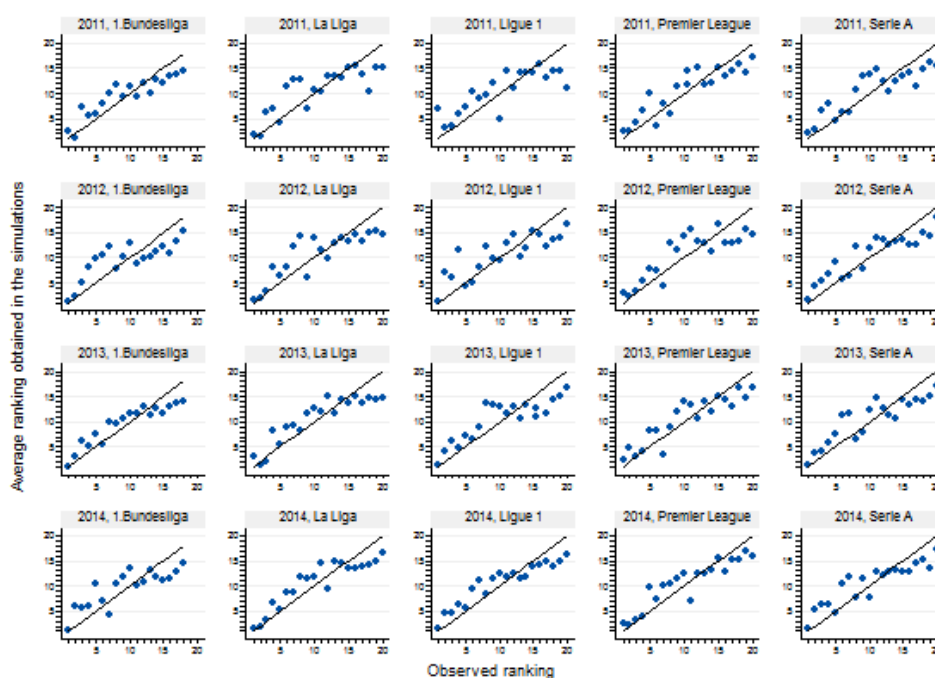


Figure 3.4: Goodness of fit : seasons 2011/2012 to 2014/2015



### 3.5.2 Out of sample predictions

Another way to test our coefficients is to use out of sample predictions. Contrary to in-sample predictions, out of sample predictions are not sensitive to overfitting (Hastie et al (2001)).

#### Standard out of sample predictions

We randomly split the sample of games in ten folds of equal size (K1, K2, ..., K10). Every fold will be used successively as a test sample.

K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
----	----	----	----	----	----	----	----	----	-----

- We estimate the coefficients on the folds K2, K3, K4, K5, K6, K7, K8, K9 and K10. We predict the probabilities of win/draw/loss (cf section 3.5.1) for the games of K1. These probabilities are out of sample predictions, as K1 was not used in the estimation.

K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
----	----	----	----	----	----	----	----	----	-----

- We estimate the coefficients on the folds K1, K3, K4, K5, K6, K7, K8, K9 and K10. We predict the probabilities of win/draw/loss for the games of K2. These probabilities are out of sample predictions, as K2 was not used in the estimation.

K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
----	----	----	----	----	----	----	----	----	-----

- ... And so on, for every fold

We end up with out of sample predictions for every game of the sample. With these probabilities and the points reward system<sup>22</sup>, we can compute the expected number of points of teams for every game :

$$E(\text{points}_{I_J}) = 3 \times Pr(I \text{ wins}) + 1 \times Pr(I \text{ draws}) + 0 \times Pr(I \text{ loses})$$

$$E(\text{points}_{J_I}) = 3 \times Pr(J \text{ wins}) + 1 \times Pr(J \text{ draws}) + 0 \times Pr(J \text{ loses})$$

Then, we can compare this expected number of points to the observed number of points in order to assess the quality of our out of sample predictions. The difference (in absolute value) between the observed number of points and the expected number of points reflects the prediction error and is an indicator of the quality of our out of sample predictions.

Our average prediction error is 1.07 points. Having a pretty large prediction error is inherent to the “*glorious uncertainty of sport*” because any prediction system will have a hard time in predicting the victory of underdogs against favorites. Therefore, in the case of sport, the prediction error must not be assessed in absolute terms but rather compared across different forecasting systems.

Betting operators use forecasting methods in order to fix their odds. Thus we can compare our out of sample predictions to the ones obtained according to the pre-game odds of the betting operators. We collected data on the odds (win/draw/loss) used by four betting operators for all the games of the five domestic leagues from 2007/2008 to 2014/2015. From the odds, we can compute the underlying probabilities used by the operators in their forecasting models. A slight adjustment needs to be made to take into account the operators’ margin.

$$Pr(I \text{ wins})_{operator} = \frac{1}{\frac{1}{odds I \text{ wins}} + \frac{1}{odds I \text{ draws}} + \frac{1}{odds I \text{ loses}}} \times \frac{1}{odds I \text{ wins}}$$

$$Pr(I \text{ draws})_{operator} = \frac{1}{\frac{1}{odds I \text{ wins}} + \frac{1}{odds I \text{ draws}} + \frac{1}{odds I \text{ loses}}} \times \frac{1}{odds I \text{ draws}}$$

$$Pr(I \text{ loses})_{operator} = \frac{1}{\frac{1}{odds I \text{ wins}} + \frac{1}{odds I \text{ draws}} + \frac{1}{odds I \text{ loses}}} \times \frac{1}{odds I \text{ loses}}$$

---

<sup>22</sup>3 points for a win, 1 point for a draw and 0 point for a loss.



The first term of these equations is used to rescale probabilities so that they sum up to one. With this rescaling, the probabilities are netted out of the operator's margin.<sup>23</sup>. From these probabilities, we can compute the expected number of points according to the different operators, the same way we did with our estimated probabilities. The difference between the observed number of points and the expected number of points derived from the operator's probabilities corresponds to the operator's prediction error. We can compare the average prediction error of the different operators to our average prediction error (table 3.12).

Our estimation	Bet365	BetWin	Interwetten	Ladbrokes
1.07 pts	1.04 pts	1.04 pts	1.05 pts	1.04 pts

Table 3.12: Average prediction error (in number of points per team)

Our average prediction error is very close to the ones derived from the operators' betting odds. This is rather impressive given the simplicity of our structural equation. In comparison with the massive amount of information that betting operators use to fix their odds, our probabilities only rely on the lineups and on the home advantage.

### Out of sample predictions on the last games of the sample

Comparing our out of sample prediction error to the ones derived from the betting operators' odds on the whole sample might be unfair because we use information that was not available to the operators when they fixed their odds. For example, our probabilities for a game in 2012/2013 will depend on games from 2014/2015 whereas the operator could not use such information when fixing its odds. A fairer way of comparing the prediction error is to do out of sample predictions on the last games of the season 2014/2015. We fix the 21<sup>st</sup> of May 2015 as a threshold date. The games that took place before this date are used to estimate the coefficients (training sample) while the game that took place after this date are not used in the estimation (test sample).



By doing so, we ensure that we do not have information on more games than the betting operators when we make our predictions. Thus the prediction errors are perfectly comparable.

<sup>23</sup>This rescaling is very important because the average margins of the operators are about 6-7 percent.  

$$Margin = 100 * \left( \frac{1}{odds\ I\ wins} + \frac{1}{odds\ I\ draws} + \frac{1}{odds\ I\ loses} - 1 \right)$$

Table 3.13 reports the average prediction error in the last games of the season 2014/2015.

Our estimation	Bet365	BetWin	Interwetten	Ladbrokes
1.14 pts	1.08 pts	1.09 pts	1.09 pts	1.09 pts

Table 3.13: Average prediction error (in number of points per team)

The prediction error is slightly higher than when we considered the whole sample. This is due to more surprising results in the end of the season. Our prediction error is very close to the ones derived from the betting operators. This shows that our coefficients are very robust : with a simple adjusted plus-minus model where players productivity do not vary over time, we end up with a prediction accuracy close to professional forecasters.

## 3.6 Extensions

### 3.6.1 Introducing time-varying productivity

One limit of our approach is that our coefficients of individual productivity do not vary over time. However, it is well established that productivity heavily depends on age in sports (see Berry, Reese & Larkey (1999)). The typical ageing curve of athletes is usually thought of as an inverted U shape: a player's productivity rises until the player reaches his highest level and then declines progressively. Therefore, the progression of a player could be modelled as a quadratic function of his age. In our case:

$$\begin{aligned}\theta_{it} &= \omega_i + \omega_{1i} \times age_{it} + \omega_{2i} \times age_{it}^2 \\ \delta_{it} &= \psi_i + \psi_{1i} \times age_{it} + \psi_{2i} \times age_{it}^2\end{aligned}$$

We allow for player specific ageing curves through the coefficients  $\omega_{1i}$ ,  $\omega_{2i}$ ,  $\psi_{1i}$  and  $\psi_{2i}$ . We could reestimate the model with these new parameters and use a penalization for the ageing curve coefficients that do not have the expected sign<sup>24</sup>. This approach might yield interesting results, but is not feasible for all players: players who are observed for one season will have both their  $\omega_{1i}$  and  $\omega_{2i}$  set to zero and players observed for exactly two seasons will have their  $\omega_{2i}$  set to zero. The rest of the players will be endowed with three times as many parameters, making the estimation procedure computationally burdensome.

---

<sup>24</sup>Due to the inverted U shape ageing curve, we expect  $\omega_{1i}$  and  $\psi_{1i}$  to be positive and  $\omega_{2i}$  and  $\psi_{2i}$  to be negative.

### 3.6.2 Application of our estimation procedure to other team settings

Our approach is not circumscribed to sports analytics. It could be adapted to other team settings where the common output is observed while individual productivity is difficult to assess. Such settings are very common as organizing work in teams has progressively become the linchpin in most organizations. For example, consider a fast food restaurant where workers strongly depend on each other (e.g. the cashier can deliver the burger if and only if another worker has prepared it) and where tasks are not fixed over time (e.g. sometimes a worker works as a cashier and sometimes he prepares the food). In this setting, we could use our fixed-effect approach to estimate individual productivity. Our structural equation (3.1) would simplify to :

$$y_I = \sum_{i \in I} \theta_i + \gamma \times Controls + \epsilon_I \quad (3.4)$$

Where

An observation is a sequence during which there is no shift among workers

$y_I$  is the collective output normalized to one unit of time (e.g. number of clients served per hour)

$i \in I$  if worker  $i$  works during the sequence considered

$\theta_i$  is the individual productivity of worker  $i$ , i.e. the parameter we are interested in

*Controls* is a set of control variables (e.g. time of the day to control for customers' attendance)

Equation (3.4) is more general than equation (3.1) because the opposition term disappeared. Moreover, in this general framework, we do not restrict  $y_I$  to be a discrete variable. It can also be a quantitative output.

The general idea of equation (3.4) is simple. In order to recover individual productivity, we look at the variations of the collective output with a lot of different combinations of workers. If workers' shifts are frequent enough and if they stay long enough in the firm<sup>25</sup>, the coefficients can be easily estimated. In order to reduce the colinearity issue due to the fact that some workers may often work with the same colleagues, a Ridge penalization could be very useful, just as in our estimation of football players' productivity. In the case where  $y_I$  is a quantitative variable, we could write the optimization problem as :

---

<sup>25</sup>Considering the fast-food restaurant example, the first condition is met while the second is more questionable.

$$\text{Min} \left( \sum_I (y_I - \sum_{i \in I} \theta_i - \gamma \times \text{Controls})^2 + \lambda \sum_z \theta_z^2 \right) \quad (3.5)$$

Thus, given a sufficient variability of the workers' lineups, it is possible to estimate individual productivity in team settings with fixed effects and shrinkage.

### 3.7 Conclusion

In this paper, we estimate football players productivity by combining a fixed-effect approach with a Ridge penalization. The fixed-effect approach allows us to consider all aspects of productivity while the Ridge penalization allows us to overcome the problem of isolating individual productivity when players often play with the same teammates. This estimation procedure could be easily adapted to other team settings where workers' shifts between teams are frequent.

Thanks to this innovative measure of individual productivity, we can test for racial discrimination in the football market at the player level. We confront our measure to players' market value in order to identify undervalued and overvalued players and we show that the share of black players is significantly higher among undervalued players than among overvalued players.

## Appendices

### Appendix A: Ordered probit

Let  $\alpha_1 < \alpha_2 < \dots < \alpha_K$  be unknown cut points to be estimated.  $K$  is the maximum number of goals scored by a team during one game

$$\begin{aligned} y_{IJ} &= 0 \text{ if } y_{IJ}^* \leq \alpha_1 \\ y_{IJ} &= k \text{ if } \alpha_k < y_{IJ}^* \leq \alpha_{k+1}, \forall k \in (1, 2, \dots, K-1) \\ y_{IJ} &= K \text{ if } y_{IJ}^* > \alpha_K \end{aligned}$$

Given the standard normal assumption for  $\epsilon_{ij}$ , we have:

$$\begin{aligned} Pr(y_{IJ} = 0) &= Pr(y_{IJ}^* \leq \alpha_1) \\ &= Pr\left(\sum_{i \in I_J} t_{i_{IJ}} \times \theta_i - \sum_{j \in J_I} t_{j_{JI}} \times \delta_j + \beta_{Home} \times Home_I + \epsilon_{IJ} \leq \alpha_1\right) \\ &= Pr(\epsilon_{IJ} \leq \alpha_1 - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I) \\ &= \Phi\left(\alpha_1 - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I\right) \end{aligned}$$

$$\begin{aligned} Pr(y_{IJ} = k) &= Pr(\alpha_k < y_{IJ}^* \leq \alpha_{k+1}) \\ &= Pr\left(\alpha_k < \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i - \sum_{j \in J_I} t_{j_{JI}} \times \delta_j + \beta_{Home} \times Home_I + \epsilon_{IJ} \leq \alpha_{k+1}\right) \\ &= Pr\left(\alpha_k - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I < \epsilon_{IJ} < \alpha_{k+1} - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I\right) \\ &= \Phi\left(\alpha_{k+1} - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I\right) - \Phi\left(\alpha_k - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I\right) \end{aligned}$$

$$\begin{aligned} Pr(y_{IJ} = K) &= Pr(y_{IJ}^* > \alpha_K) \\ &= 1 - Pr(y_{IJ}^* \leq \alpha_K) \\ &= 1 - Pr\left(\sum_{i \in I_J} t_{i_{IJ}} \times \theta_i - \sum_{j \in J_I} t_{j_{JI}} \times \delta_j + \beta_{Home} \times Home_I + \epsilon_{IJ} \leq \alpha_K\right) \\ &= 1 - Pr\left(\epsilon_{IJ} \leq \alpha_K - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I\right) \\ &= 1 - \Phi\left(\alpha_K - \sum_{i \in I_J} t_{i_{IJ}} \times \theta_i + \sum_{j \in J_I} t_{j_{JI}} \times \delta_j - \beta_{Home} \times Home_I\right) \end{aligned}$$

where  $\Phi$  is the cdf of the standard Normal.

## Appendix B: Productivity ranking (top 100 - all positions)

Ranking	Player	ME_total	ME_offensive	ME_defensive
1	Lionel Messi	.351	.351	
2	Cristiano Ronaldo	.333	.333	
3	Cesc Fabregas	.302	.243	.059
4	Frank Lampard	.274	.205	.069
5	Yaya Toure	.255	.138	.117
6	Karim Benzema	.253	.253	
7	Bastian Schweinsteiger	.246	.171	.075
8	Xavi	.241	.175	.066
9	Arturo Vidal	.238	.128	.11
10	Franck Ribery	.237	.237	
11	Gonzalo Higuain	.235	.235	
12	Mesut Ozil	.228	.228	
13	Andres Iniesta	.224	.13	.094
14	Luka Modric	.218	.125	.093
15	Arjen Robben	.217	.217	
16	Kwadwo Asamoah	.194	.103	.091
17	Wayne Rooney	.186	.186	
18	Marco Verratti	.183	.118	.065
19	Thomas Muller	.182	.182	
20	Zlatan Ibrahimovic	.181	.181	
21	Eden Hazard	.181	.181	
22	Luis Suarez	.179	.179	
23	Javier Pastore	.178	.178	
24	Ivan Rakitic	.176	.087	.089
25	Pedro	.176	.176	
26	Robert Lewandowski	.174	.174	
27	Paul Scholes	.171	.125	.045
28	Edin Dzeko	.169	.169	
29	Dani Parejo	.165	.133	.033
30	John Terry	.165		.165
31	Rafinha	.165		.165
32	Alexis Sanchez	.164	.164	
33	Tiago	.163	.091	.071
34	Darren Fletcher	.162	.108	.054
35	Sergio Aguero	.159	.159	

Ranking	Player	ME_total	ME_offensive	ME_defensive
36	Mario Gotze	.156	.156	
37	Asier Illarramendi	.155	.112	.043
38	Sulley Muntari	.155	.066	.089
39	Nemanja Vidic	.153		.153
40	Nuri Sahin	.149	.097	.053
41	Giorgio Chiellini	.149		.149
42	Kaka	.148	.148	
43	Francesco Totti	.145	.145	
44	Mario Gomez	.145	.145	
45	Yohan Cabaye	.143	.079	.063
46	Thiago	.141	.117	.023
47	Medhi Benatia	.14		.14
48	Juanfran	.139		.139
49	Jeremy Toulalan	.139		.139
50	Ashley Young	.139	.139	
51	Gareth Bale	.139	.139	
52	Aaron Ramsey	.136	.063	.073
53	Darron Gibson	.135	.094	.041
54	Jesus Navas	.134	.134	
55	Angel Di Maria	.134	.134	
56	Gonzalo Castro	.133	.134	-.002
57	Dante	.132		.132
58	Dejan Stankovic	.131	.101	.031
59	Nani	.13	.13	
60	Jakub Blaszczykowski	.13	.13	
61	Carlos Tevez	.128	.128	
62	Seydou Keita	.127	.073	.055
63	Moussa Dembele	.127	.078	.049
64	Rodrigo Taddei	.126	.078	.048
65	David Villa	.125	.125	
66	Paul Pogba	.125	.052	.074
67	Michael Ballack	.124	.089	.035
68	Mirko Vucinic	.124	.124	
69	Daniel Sturridge	.123	.123	
70	Marek Hamsik	.123	.123	

Ranking	Player	ME_total	ME_offensive	ME_defensive
71	Claudio Marchisio	.123	.077	.045
72	Alberto Aquilani	.123	.085	.037
73	Robin van Persie	.122	.122	
74	Xabi Alonso	.121		.121
75	Gerard Pique	.121		.121
76	Lucho Gonzalez	.12	.103	.016
77	Claudio Pizarro	.12	.12	
78	Samir Nasri	.12	.12	
79	Jerome Boateng	.12		.12
80	Maxwell	.118		.118
81	Alvaro Arbeloa	.118		.118
82	Didier Drogba	.117	.117	
83	Jamie Carragher	.117		.117
84	Fernando	.116	.088	.028
85	Goran Pandev	.116	.116	
86	Ryan Giggs	.115	.054	.06
87	Ezequiel Lavezzi	.115	.115	
88	Thiago Silva	.114		.114
89	Robinho	.114	.114	
90	Juan Mata	.113	.113	
91	Theo Walcott	.112	.112	
92	Dimitri Payet	.111	.111	
93	Hernanes	.111	.034	.077
94	Philipp Lahm	.111		.111
95	Javier Mascherano	.111		.111
96	Miranda	.11		.11
97	Ever Banega	.11	.081	.03
98	Gabi	.11	.042	.067
99	Yossi Benayoun	.108	.108	
100	Loïc Perrin	.108		.108

Seasons 2007/2008 to 2014/2015. Lambda = 200. Only midfielders have both an offensive and a defensive contribution.





# Bibliography

- Alchian, A. A. & Demsetz, H. (1972), ‘Production, Information Costs, and Economic Organization’, *American Economic Review* **62**(5), 777–795.
- Altman, D. (2013), ‘Introducing the shapley value to football’, *Stats Insights* .
- Apestequia, J. & Palacios-Huerta, I. (2010), ‘Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment’, *American Economic Review* **100**(5), 2548–2564.
- Ariely, D., Gneezy, U., Loewenstein, G. & Mazar, N. (2009), ‘Large Stakes and Big Mistakes’, *Review of Economic Studies* **76**(2), 451–469.
- Azmat, G., Calsamiglia, C. & Iriberry, N. (2016), ‘Gender differences in response to big stakes’, *Journal of the European Economic Association* **14**(6), 1372–1400.
- Berri, D. J. (1999), ‘Who is ‘most valuable’? measuring the player’s production of wins in the national basketball association’, *Managerial and decision economics* pp. 411–427.
- Berri, D. & Schmidt, M. (2010), *Stumbling on wins in basketball*, Pearson Education.
- Berry, S. M., Reese, C. S. & Larkey, P. D. (1999), ‘Bridging different eras in sports’, *Journal of the American Statistical Association* **94**(447), 661–676.
- Betts, J. R. et al. (2011), ‘The economics of tracking in education’, *Handbook of the Economics of Education* **3**(341-381), 4.
- Bilger, M., Manning, W. G. et al. (2011), ‘Measuring overfitting and misspecification in nonlinear models’, *Health, Econometrics and Data Group Working Paper* **11**, 25.
- Black, D. A., Kolesnikova, N., Sanders, S. G. & Taylor, L. J. (2013), ‘The role of location in evaluating racial wage disparity’, *IZA Journal of Labor Economics* **2**(1), 2.

- Blau, F. D. & Beller, A. H. (1992), 'Black-white earnings over the 1970s and 1980s: Gender differences in trends', *The Review of Economics and Statistics* pp. 276–286.
- Breda, T. & Hillion, M. (2016), 'Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in france', *Science* **353**(6298), 474–478.
- Breda, T. & Ly, S. T. (2015), 'Professors in core science fields are not always biased against women: Evidence from france', *American Economic Journal: Applied Economics* **7**(4), 53–75.
- Brown, J. (2011), 'Quitters never win: The (adverse) incentive effects of competing with superstars', *Journal of Political Economy* **119**(5), 982–1013.
- Buchanan, J. M., Tollison, R. D. & Tullock, G. (1980), *Toward a Theory of the Rent-seeking Society*, number 4, Texas A & M Univ Pr.
- Caille, J.-P. (2007), 'Perception du système éducatif et projets d'avenir des enfants d'immigrés', *Éducation et formations* **74**, 117–142.
- Cain, G. G. (1986), 'The economic analysis of labor market discrimination: A survey', *Handbook of labor economics* **1**, 693–785.
- Carneiro, P., Heckman, J. J. & Masterov, D. V. (2005), 'Labor market discrimination and racial differences in premarket factors', *The Journal of Law and Economics* **48**(1), 1–39.
- Charness, G. & Dufwenberg, M. (2006), 'Promises and Partnership', *Econometrica* **74**(6), 1579–1601.
- Chen, H. & Lim, N. (2013), 'Should Managers Use Team-based Contests?', *Management Science* **59**(12), 2823–2836.
- Corcoran, M. & Duncan, G. J. (1979), 'Work history, labor force attachment, and earnings differences between the races and sexes', *Journal of Human Resources* pp. 3–20.
- Darley, J. M. & Gross, P. H. (1983), 'A hypothesis-confirming bias in labeling effects.', *Journal of Personality and Social Psychology* **44**(1), 20.
- Davaillon, A. & Nauze-Fichet, E. (2004), 'Les trajectoires scolaires des enfants pauvres', *Éducation et formations* **70**, 41.

- Dechenaux, E., Kovenock, D. & Sheremeta, R. M. (2015), ‘A Survey of Experimental Research on Contests, All-pay Auctions and Tournaments’, *Experimental Economics* **18**(4), 609–669.
- Deschamps, P. & De Sousa, J. (2014), ‘Labor mobility and racial discrimination’, *LIEPP Working Paper* **35**.
- Duru-Bellat, M. (1990), *L'école des filles: quelle formation pour quels rôles sociaux?*, l'Harmattan Paris.
- Duru-Bellat, M., Jarousse, J.-P. & Mingat, A. (1993), ‘Les scolarités de la maternelle au lycée: étapes et processus dans la production des inégalités sociales’, *Revue française de sociologie* pp. 43–60.
- Elhoweris, H., Mutua, K., Alsheikh, N. & Holloway, P. (2005), ‘Effect of children’s ethnicity on teachers’ referral and recommendation decisions in gifted and talented programs’, *Remedial and Special Education* **26**(1), 25–31.
- Franck, E. & Nüesch, S. (2010), ‘The effect of talent disparity on team productivity in soccer’, *Journal of Economic Psychology* **31**(2), 218–229.
- Friedman, J., Hastie, T. & Tibshirani, R. (2001), *The elements of statistical learning*, Vol. 1, Springer series in statistics New York.
- Fu, Q., Ke, C. & Tan, F. (2015), ‘“Success Breeds Success” or “Pride Goes Before a Fall”: Teams and Individuals in Multi-Contest Tournaments’, *Games and Economic Behavior* **94**, 57–79.
- Fu, Q., Lu, J. & Pan, Y. (2015), ‘Team Contests with Multiple Pairwise Battles’, *American Economic Review* **105**(7), 2120–40.
- Glock, S., Krolak-Schwerdt, S., Klapproth, F. & Böhmer, M. (2013), ‘Beyond judgment bias: How students’ ethnicity and academic profile consistency influence teachers’ tracking judgments’, *Social Psychology of Education* **16**(4), 555–573.
- Goux, D. & Maurin, E. (2007), ‘Close neighbours matter: Neighbourhood effects on early performance at school’, *The Economic Journal* **117**(523), 1193–1215.
- Gramacy, R. B., Jensen, S. T. & Taddy, M. (2013), ‘Estimating player contribution in hockey with regularized logistic regression’, *Journal of Quantitative Analysis in Sports* **9**(1), 97–111.

- Guryan, J., Kroft, K. & Notowidigdo, M. J. (2009), ‘Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments’, *American Economic Journal: Applied Economics* **1**(4), 34–68.
- Guyon, N. & Huillery, E. (2014), ‘Choix d’orientation et origine sociale: mesurer et comprendre l’autocensure scolaire’, *LIEPP Report* **3**.
- Guyon, N. & Huillery, E. (2016), ‘Biased aspirations and social inequality at school: Evidence from french teenagers’, *Sciences Po Working Paper* .
- Hamilton, B. H., Nickerson, J. A. & Owan, H. (2003), ‘Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation’, *Journal of political Economy* **111**(3), 465–497.
- Hanna, R. N. & Linden, L. L. (2012), ‘Discrimination in grading’, *American Economic Journal: Economic Policy* **4**(4), 146–168.
- Harris, C. & Vickers, J. (1987), ‘Racing with Uncertainty’, *Review of Economic Studies* **54**(1), 1–21.
- Hoerl, A. E. & Kennard, R. W. (1970), ‘Ridge regression: Biased estimation for nonorthogonal problems’, *Technometrics* **12**(1), 55–67.
- Hoxby, C. & Avery, C. (2013), ‘The missing one-offs: The hidden supply of high-achieving, low-income students’, *Brookings Papers on Economic Activity* **2013**(1), 1–65.
- Huang, L. (2016), ‘Is there no ‘I’ in Team? Rational Players in Dynamic Team Competition’, *Working paper. Available at SSRN: <https://ssrn.com/abstract=2879863>* .
- Huang, L. & Murad, Z. (2017), ‘Fighting Alone or Fighting for a Team: Evidence from Experimental Pairwise Contests’, *Better for Less, Discussion Paper Series* (06/2016).
- Ilardi, S. & Barzilai, A. (2008), ‘Adjusted plus-minus ratings: New and improved for 2007-2008’, <http://www.82games.com/ilardi2.htm> .
- Kandel, E. & Lazear, E. P. (1992), ‘Peer Pressure and Partnerships’, *Journal of Political Economy* **100**(4), 801–817.

- Klumpp, T. & Polborn, M. K. (2006), ‘Primaries and the New Hampshire Effect’, *Journal of Public Economics* **90**(6), 1073–1114.
- Knoeber, C. R. & Thurman, W. N. (1994), ‘Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production’, *Journal of Labor Economics* **12**(2), 155–179.
- Kocher, M. G., Lenz, M. V. & Sutter, M. (2012), ‘Psychological Pressure in Competitive Environments: New Evidence from Randomized Natural Experiments’, *Management Science* **58**(8), 1585–1591.
- Konrad, K. A. & Kovenock, D. (2009), ‘Multi-battle Contests’, *Games and Economic Behavior* **66**(1), 256–274.
- Kuper, S. (2014), *Soccernomics: Why England Loses, Why Spain, Germany, and Brazil Win, and Why the US, Japan, Australia and Even Iraq Are Destined to Become the Kings of the World’s Most Popular Sport*, Nation Books.
- Lang, K. & Lehmann, J.-Y. K. (2012), ‘Racial discrimination in the labor market: Theory and empirics’, *Journal of Economic Literature* **50**(4), 959–1006.
- Lang, K. & Manove, M. (2011), ‘Education and labor market discrimination’, *American Economic Review* **101**(4), 1467–96.
- Lavy, V. (2008), ‘Do gender stereotypes reduce girls’ or boys’ human capital outcomes? evidence from a natural experiment’, *Journal of public Economics* **92**(10), 2083–2105.
- Lazear, E. P. & Rosen, S. (1981), ‘Rank-order Tournaments as Optimum Labor Contracts’, *Journal of Political Economy* **89**(5), 841–864.
- Mago, S. D., Sheremeta, R. M. & Yates, A. (2013), ‘Best-of-three contest experiments: Strategic versus psychological momentum’, *International Journal of Industrial Organization* **31**(3), 287–296.
- Malueg, D. A. & Yates, A. J. (2010), ‘Testing Contest Theory: Evidence from best-of-three Tennis Matches’, *Review of Economics and Statistics* **92**(3), 689–692.
- Mas, A. & Moretti, E. (2009), ‘Peers at Work’, *American Economic Review* **99**(1), 112–145.
- Merle, P. (2014), ‘Faut-il en finir avec les notes?’, *La vie des idées* **2**.

- Müller, J., Upmann, T. & Prinz, J. (2013), ‘Individual team productivity-a conceptual approach’, *Tinbergen Institute Discussion Paper 13-183/I* .
- Neal, D. A. & Johnson, W. R. (1996), ‘The role of premarket factors in black-white wage differences’, *Journal of political Economy* **104**(5), 869–895.
- Oaxaca, R. L. & Ransom, M. R. (1994), ‘On discrimination and the decomposition of wage differentials’, *Journal of econometrics* **61**(1), 5–21.
- O’Neill, J. (1990), ‘The role of human capital in earnings differences between black and white men’, *Journal of economic Perspectives* **4**(4), 25–45.
- Prendergast, C. (1999), ‘The Provision of Incentives in Firms’, *Journal of Economic Literature* **37**(1), 7–63.
- Rasch, G. (1993), *Probabilistic models for some intelligence and attainment tests.*, ERIC.
- Reimers, C. W. (1983), ‘Labor market discrimination against hispanic and black men’, *The review of economics and statistics* pp. 570–579.
- Rodgers, W. M. & Spriggs, W. E. (1996), ‘The effect of federal contractor status on racial differences in establishment-level employment shares: 1979-1992’, *The American Economic Review* **86**(2), 290–293.
- Rosenbaum, D. T. (2004), ‘Measuring how nba players help their teams win’, *82Games.com* (<http://www.82games.com/comm30.htm>) pp. 4–30.
- Rosenqvist, O. & Skans, O. N. (2015), ‘Confidence Enhanced Performance? – The Causal Effects of Success on Future Performance in Professional Golf Tournaments’, *Journal of Economic Behavior & Organization* **117**, 281–295.
- Sheremeta, R. M. (2017), ‘Behavior in Group Contests: A Review of Experimental Research’, *Journal of Economic Surveys* .
- Smith, J. P. & Welch, F. R. (1986), *Closing the Gap: Forty Years of Economic Progress for Blacks.*, ERIC.
- Sprietsma, M. (2013), ‘Discrimination in grading: experimental evidence from primary school teachers’, *Empirical Economics* pp. 1–16.

- Szymanski, S. (2000), 'A market test for discrimination in the english professional soccer leagues', *Journal of political Economy* **108**(3), 590–603.
- Szymanski, S. (2003), 'The assessment: the economics of sport', *Oxford Review of Economic Policy* **19**(4), 467–477.
- Tenenbaum, H. R. & Ruck, M. D. (2007), 'Are teachers' expectations different for racial minority than for european american students? a meta-analysis.', *American Psychological Association* .
- Thomas, A., Ventura, S. L., Jensen, S. T. & Ma, S. (2013), 'Competing process hazard function models for player ratings in ice hockey', *The Annals of Applied Statistics* pp. 1497–1524.
- Tiedemann, T., Francksen, T. & Latacz-Lohmann, U. (2011), 'Assessing the performance of german bundesliga football players: a non-parametric metafrontier approach', *Central European Journal of Operations Research* **19**(4), 571–587.
- Van Ewijk, R. (2011), 'Same work, lower grade? student ethnicity and teachers' subjective assessments', *Economics of Education Review* **30**(5), 1045–1058.
- Vrignaud, P. (2016), 'L'évolution des intentions d'orientation et du choix professionnel au cours du college : l'impact du genre et de l'origine sociale', *CNESCO report* .