Do Star Performers Produce More Stars?
Peer Effects and Learning in Elite Teams

August 2014

ABSTRACT

This study investigates the professional soccer industry to ask whether the talent of an individual's co-workers helps explain differences in the rate of human capital accumulation on the job. Data tracking national soccer team performance and the professional leagues their members play for are particularly well suited for developing convincing non-experimental evidence about these kinds of peer effects. The empirical results consistently show that performance improves more after an individual has been a member of an elite team than when he has been a member of lower level teams. The conclusion is borne out by a rich set of complementary data on: national team performance, player-level performance, performance of foreign players who joined elite teams after an exogenous shift in the number of foreign players participating on top club teams, performance of players on national teams in the year just before and the year just after they join an elite club team, and experiences of several national team players obtained through personal interviews.

Casey Ichniowski
Graduate School of Business
and NBER
713 Uris Hall
Columbia University
New York, NY 10027

Anne Preston
Department of Economics
203 Stokes Hall
Haverford College
Haverford, PA 19041
apreston@haverford.edu

The authors thank Michael Egley, Damien Fenske-Corbiere, Kazunari Inoki, Amy Johnson, Jack Koehler, and Frankie Pavia for research assistance on this project and Raymond Lim for expert work on the data set construction and analysis. We thank participants at the Haverford-Swarthmore and Columbia Business School seminars for helpful comments. We thank Bob Gibbons, Florian Ederer, and Alan Benson for comments on earlier drafts. We also offer special thanks to several members of the U.S. Men's National Team for agreeing to interviews for this study and to Sunil Gulati for arranging these interviews.

I. Introduction

Recent empirical research analyzing micro-level data sets has provided new insights into the nature of real firm-level production functions. This "insider" research on specific industries and production processes identifies a rich set of factors and dynamics that determine a firm's productivity and therefore suggests promising avenues for developing a more realistic theory of the firm.[1] One consistent theme from this emerging literature is that peer effects matter. A worker's productivity is affected in interesting ways simply by the fact that she has co-workers.

Peer effects impact worker productivity in different ways. Recent studies suggest two conclusions. First, high performing workers are important. Productivity effects of higher performing workers on lower productivity co-workers have been documented for shift workers (Mas and Moretti, 2009), work group managers (Lazear, Shaw, and Stanton, 2012), and individuals in experimental settings (Falk and Ichino, 2006). Second, peer effects in these studies are temporary and generally due to monitoring by the higher performing worker, direct supervision by better managers, or the proximity of co-workers in experimental tasks.[2] When the high performing worker or manager is no longer present, the productivity of the peer declines. In this study, we consider the possibility of more permanent and transferable performance increases due to high performing co-workers. To address the many challenges of estimating peer productivity effects, we follow the approach of collecting detailed data on one industry setting and production process. These micro-level data are needed to identify who works with whom in group settings and to measure group and worker performance in convincing ways. Furthermore, the setting for our study has several unique features that allow us to develop especially convincing estimates of the effects of peers on worker performance.

Specifically, we assemble rich data on the performance of soccer teams and individual soccer players that let us test for these peer effects. First, because top flight soccer players can play on two teams – their country's national team and professional club teams – we can readily address the reflection problem not just by measuring the performance effects of players who switch teams (Ichino and Maggi, 2000), but by measuring a player's exposure to an elite club team and then tracking his performance when he returns to play for his national team. Second, a

[1] For a review, see Ichniowski and Shaw (2012).

[2] Some studies on peer effects in academic settings (Sacerdote 2001, Imberman, Kugler and Acerdote 2012) and in program participation (Dahl, Loken and Mogstad, 2014) show longer lasting effects, although none that are permanent.

unique feature of the institutional setting of professional soccer in Europe is the existence of a legal change – the landmark Bosman ruling – that led to a dramatic increase in the number of foreign players that could play in elite European leagues, thus allowing us to use a regression discontinuity design and make stronger causal inferences about any estimated peer effects. Third, we can account for selectivity effects of better players joining top clubs in a second way to help isolate the effects of learning from high talent peers. In particular, we examine the contribution of national team players who join elite clubs in the year just before and the year just after they join an elite club for the first time. Finally, we have collected very rich data on "events" in a large number of soccer games and use these data to create detailed information on player-level performance. We analyze whether the patterns that we document in models of team-level performance are supported by similar patterns on individual level performance in a different data set with a different level of analysis. This complement of research design features is unique and permits the estimation of especially convincing models of longer-lasting productivity effects that are transferable across team settings and are due to an individual's exposure to high performing peers.

The study is organized as follows. Section II identifies the type of group learning parameter we are estimating in the empirical work and considers the theoretical importance of longer-lasting peer effects. Section III describes the team-level data sets and samples we have constructed for this study. Sections IV through VII present empirical results on peer effects estimated from team-level panel data, including estimates from fuzzy regression discontinuity models that focus on the period after the Bosman ruling caused an exogenous shift in the number of foreign players in top leagues. Sections VIII and IX present player-level data and results concerning peer effects that players enjoy when they join elite club teams. Section X reports evidence from interviews with several national team players and section XI concludes.

II. Long-Lasting Transferable Peer Effects in Elite Teams

A number of insider studies on specific industries have documented important productivity effects due to peers. Here we briefly review several recent studies on this topic. We then present a simple model that illustrates the interesting implications of more permanent peer productivity effects that last longer than the time that co-workers and managers monitor their peers and which are portable to similar production settings elsewhere.

<u>Empirical Studies of Productivity Effects due to Peers</u>

Several recent studies investigate productivity spillovers across peer workers. One theme from these studies is high performing workers appear to be especially important. Bringing a new high-performing worker onto a shift changes the productivity of co-workers (Mas and Moretti, 2009). Good managers can consistently increase productivity as they move from one work group to the next (Lazear, Shaw, Stanton, 2012). This finding about the positive effects of higher performing workers has also been observed in controlled laboratory settings (Falk and Ichino, 2006). The productivity effects due to peers documented in these studies appear to be temporary effects. When better performing shift workers are not observing lower performers, or when the better manager is no longer supervising the group, performance declines.

Related empirical evidence is available in studies of how student performance is affected by the nature of his or her classmates, and in particular whether “tracking” of students into higher and lower ability groups will impact performance. With careful methodical approaches and comparisons, a few studies on this topic offer somewhat mixed results. Sacerdote (2001) takes advantage of random pairing of roommates in first year rooming groups at Dartmouth College to estimate positive effects of a roommate’s freshman GPA on own GPA. Duflo, Dupas, and Kremer (2011) investigate how random assignment of elementary school children in India into tracked versus not tracked settings impacts their performance. They document positive effects of tracking, and focus on potential benefits of students benefiting from high ability peers in upper tracks and teacher’s ability to tailor their plans across all tracks. Similarly, Pop-Eleches and Urquiola (2013) use a regression discontinuity design to identify a positive effect of higher ranked schools on baccalaureate exam scores for Romanian high school students. In contrast, using a similar research methodology, Abdulkadiroglu, Angrist, and Pathak (2013) find no effect on standardized test scores of studying in “exam schools” versus other schools in Boston or New York City. Similarly Dobbie and Fryer (2014) find no effect of attending the New York City “exam schools” on college enrollment, graduation or quality.

A Model of Productivity Effects from Peers in Elite Teams

In this study, we are interested in productivity effects due to peers in work settings like the first set of studies, but we focus on longer lasting increases in human capital that cannot be explained by monitoring or supervision activities when higher performing co-workers or managers are present. How might high talent peers impact their co-workers? And if star co-workers do improve productivity of others around them in a way that lasts beyond the time when the stars are watching over their peers, what are the implications of these kinds of peer effects?

Consider individuals who are members of work teams for three periods during their careers. The quality of each individual on the firm's "home office team" is drawn from a normal distribution at the beginning of the first period. All individuals work for their own home office team in all three periods. Workers also spend time on a second team that works on similar projects but with different teammates. Some second teams are special "elite teams" while others are not. If an individual's quality is above some threshold, Q_{Thresh}, which cuts off an upper tail of the quality distribution, he works on one of these "elite teams" with other high performing individuals. Other employees spend similar amounts of time on a second team, but these teams do not have elite members.

Membership on an elite team where all individuals are high quality workers improves the quality of the members in the following year by some amount $\overline{E} > 0$. We do not dictate the method through which quality improves, but it could be through training with better individuals, opportunities to observe superior work, competing against better workers, or active mentoring by better workers. In the other non-elite second teams, individuals do not receive any bump in quality from their work in those lower quality second teams. These individuals remain at the same quality level unless some of their team mates on the home office team have worked on the elite teams. Thus, some benefits from membership on elite teams are allowed to spill over to non-elite individuals on the home office teams. We let this productivity improvement be less pronounced than improvement from being a member of an elite team, and also let it be positively related to the proportion of individuals on the home office team who are members of elite second teams.

We can summarize the quality of individuals over three periods as follows:

Q_{ij}^{1} = quality of player i on home office team j during period 1, with $Q_{ij}^{1} \sim N(\mu_Q, \sigma_Q)$

$$Q_{ij}^{2} = Q_{ij}^{1} + \max(E_{ij}^{1}, \alpha PCTE_{j}^{1}),$$

where E is the quality boost from playing on an elite team; $E_{ij}^{1} = 0$ if $Q_{ij}^{1} < Q_{Thresh}$ and $E_{ij}^{1} = \overline{E} > 0$ if $Q_{ij}^{1} \geq Q_{Thresh}$; $PCTE_{j}^{1}$ is % of the individuals on home office team j in period 1 who are members of special elite second teams players; and $0 < \alpha < \overline{E}$.

$$Q_{ij}^{3} = Q_{ij}^{2} + \max(E^{2}{}_{ij}, \alpha PCTE_{j}^{2})$$

where $E_{ij}^{2} = 0$ if $Q_{ij}^{2} < Q_{Thresh}$ and $E_{ij}^{2} = \overline{E} > 0$ if $Q_{ij}^{2} \geq Q_{Thresh}$; $PCTE_{j}^{2}$ is the percent of individuals on home office team j in period 2 who are also members of elite second teams; and $0 < \alpha < \overline{E}$.

Team Performance

When teams have such peer effects, what does average quality of the (home office) teams look like after three periods? To illustrate the answer to this question, let the home office teams employ individuals equally distributed between first, second, and third period players (i.e. a third of home team players will be first period, a third will be second period, and a third will be third period players). If home office team performance ($PERF_{j}^{t}$) depends on the average quality of its members, then we can represent team performance by:

$$PERF_{j}^{t} = f(Q_{j,avg}^{t}) \text{ where } f' > 0 \text{ and } f'' < 0$$

$$Q_{j,avg}^{t} = \sum_{i=1}^{k} Q_{ij}^{1}/3k + \sum_{i=k+1}^{2k} (Q_{ij}^{1} + \max(E_{ij}^{1}, \alpha PCTE_{j}^{t-1}))/3k + \sum_{i=2k+1}^{3k} (Q_{ij}^{1} + \max(E_{ij}^{1}, \alpha PCTE_{j}^{t-1}) + \max(E_{ij}^{2}, \alpha PCTE_{j}^{t-2}))/3k$$

Consider Team 1 that has three years with no Elite players. The average expected quality of players will be:

$Q^t_{1,avg} = \sum_{i=1}^{3k} Q^L_{i1} / 3k$ where Q^L represents quality of players below the cutoff.

$$E(Q^L \mid Q < Q_{Thresh}) = \mu_Q - \sigma_Q \left[\frac{\phi(\frac{Q_{Thresh} - \mu_Q}{\sigma_Q})}{\Phi(\frac{Q_{Thresh} - \mu_Q}{\sigma_Q})} \right] = \mu_Q - \chi$$

Therefore $E(Perf^t_1) = f(E(Q^t_{1,avg}) = f(E(Q^L) = f(\mu_Q - \chi)$

Consider Team 2 with x<k 1st period players who are elite, y<k 2nd period players who are elite, and z<k 3rd period players who are elite.

$$Q^t_{2,avg} = (\sum_{i=1}^{k-x} Q^{1,L}_{ij} + \sum_{i=1}^{x} Q^{1,H}_{ij}) / 3k +$$
$$(\sum_{i=k+1}^{2k-y} (Q^{1L}_{ij} + \alpha PCTE^{t-1}) + \sum_{i=1}^{y} (Q^{1,H}_{ij} + \bar{E})) / 3k + (\sum_{i=2k+1}^{3k-z} (Q^{1,L}_{ij} + \alpha(PCTE^{t-1} + PCTE^{t-2}) + \sum_{i=1}^{z} (Q^{1,H}_{ij} + 2\bar{E})) / 3K$$

This somewhat cumbersome expression for Q_{avg} can be simplified if we assume a steady stream of workers who qualify for the elite teams, i.e. $PCTE^t = PCTE^{t-1} = PCTE^{t-2} = PCTE$. Further we acknowledge that

$$E(Q^H \mid Q > Q_{Thresh}) = \mu_Q + \sigma_Q \left[\frac{\phi(\frac{Q_{Thresh} - \mu_Q}{\sigma_Q})}{1 - \Phi(\frac{Q_{Thresh} - \mu_Q}{\sigma_Q})} \right] = Q_{Thresh} + \lambda .$$

Then expected average quality is given by:

$$E(Q^t_{2,avg}) = (1 - PCTE)(\mu_Q - \chi + \alpha PCTE) + PCTE(Q_{Thresh} + \lambda + \bar{E})$$

And expected performance is given by :

$$E(Perf_2) = f(E(Q^t_{2,avg})) = f((1 - PCTE)(\mu_Q - \chi + \alpha PCTE) + PCTE(Q_{Thresh} + \lambda + \bar{E}))$$

Clearly the performance of a home team like team 2 that has members who participate in second teams that are elite will be higher than performance of teams like team 1 with all non-elite members, and the performance differential between the two types of teams will increase with percent elite in the second team.

Drawing a high quality individual in the first period has important effects on the performance of the home office team. First, it pulls up the average quality of the team for three periods. Second, the high quality individual gets a quality boost from his participation on elite second teams comprised solely of other high quality players, and this boost in turn spills over to the home office team members in his second and third periods with the home office team. Finally, though not shown above, it could also be the case that non-elite individuals improve performance as they work with more elite individuals on the home team, and borderline workers may possibly get enough of a boost to become elite workers.

Implications and Possible Extensions

While we develop this model of individuals working in two teams of varying quality in part to motivate the empirical work below (where the home office team is a country's national soccer team, and the elite and non-elite second teams are professional club teams of varying quality), the concept of having top employees learn from the best of the best may have broader implications. Several features of the model should be highlighted. First, the positive peer effects described above that arise when high quality workers train and compete with other high quality workers are effects that persist across other work environments without elite co-workers.[3] Therefore these are not peer effects due to monitoring or supervision but true increases in human capital. Second, because the quality improvement persists when transferring teams, there can be secondary peer effects as non-elite workers improve while working with elite workers on home teams. Finally, an exceptional draw in any period improves team performance for several periods into the future that can last for the span of the career of the cohort of high quality players.

The model highlights several interesting dynamics that longer-lasting peer effects could produce. First, team performance diverges over time. The peer effects given by $\overline{E}$ ultimately

[3] These relationships which generate positive peer effects are similar to the relationships that arise when an academic co-authors with high quality collaborators at elite institutions.

cause the upper tail of worker quality to shift further to the right, so that team performance – given by average worker quality – increases in teams where these peer effects occur. The persistence of these team-level performance differentials will of course depend on a number of factors not modeled above – including turnover and tenure of team members, opportunities for attracting team members from other teams, how severe and immediate market pressures are for eliminating lower performing teams – but the model does suggest a mechanism through which longer lasting group-level productivity differentials can be generated.

Moreover, while the model above does not consider compensation of team members, the (lucky) high performing teams will, by period 2, likely have at least one wage bill advantage over the (unlucky) lower performing teams. If the quality of the teams and the opportunities to learn in high quality teams are known, then new workers will accept lower wages to join higher quality teams since they would be gaining useful general human capital from peers in those teams. Finally, while we allow differences in initial team quality to occur through random assignment of different quality workers, the model certainly suggests that cost-effective means of identifying elite workers – possibly simply because the pool of workers for one team is better than the pool of workers available to some other team as is the case of national soccer teams that we analyze below – will have important and long-lasting effects.

III. Data and Samples for Team-Level Analyses

This study's empirical analysis of the effects of membership on an elite team on the subsequent performance of team members employs personally collected data on professional soccer teams and players. We analyze several convincing and accurate performance outcomes at two different levels of analysis – the national team-level and the player-level. At the heart of the empirical analysis is a comparison of performance outcomes for a player's national team and for the player himself in national team games when he plays on club teams of different quality. The "treatment" is exposure to elite teams and teammates, and the "response" is the change in performance after this exposure. This section describes the construction of the national team data set and samples. Player-level data are described when we introduce those models below.

National Team Panel Data Set

We collect data on two teams a professional soccer player can play on – his own country's national team and his professional club team. The basic structure of the data set is panel data measuring the performance and composition of national soccer teams. There are up to twenty-one years of data from 1990-2010 on 101 countries' national soccer teams. Central to the empirical analyses, the data set includes measures of: (a) national team performance and (b) the club affiliations of the national team members with a particular emphasis on differentiating elite versus other clubs. Other variables collected for these country-year observations include age of the players on the roster and country characteristics.

Performance Measures for National Soccer Teams

The setting of professional soccer offers especially convincing measures of performance. At the team level, we use widely accepted measures of success rates of national teams, including one that builds in adjustments for many determinants of wins and losses such as the quality of opponents and the importance of the match.

The most widely publicized rankings of national soccer teams are the Fédération Internationale de Football Association (FIFA) rankings. While we have collected and will analyze FIFA rankings for all countries, there are a number of potential limitations with FIFA rankings. First, the rankings only began in 1993 and so will not be available for three important early (pre-Bosman ruling) years in the sample. Second, since their introduction in December 1992, the rankings have changed several times. Before 1999, FIFA rankings were based on a simple 3-1-0 score for win, draw or loss. After 1999, the ranking was changed to allow differences in points earned depending on the difficulty of the opponent, the goal differential, and the importance of the match. FIFA rankings have been based on moving averages of several years' worth of win-loss records but the number of years of outcomes considered has changed over time. By the end of 1999, after the FIFA ranks had existed for eight years, a current year's FIFA ranking was calculated as a weighted average of the last eight years' worth of rankings with more recent years' rankings weighted more heavily. Prior to that, the FIFA rankings were based on fewer years since they had only been established in late 1992. After 2006, the moving average changed from an eight-year to a four-year moving average calculation, again with the more recent years weighted more heavily. Today, the current year's points get a weight of 1, and the weights for year minus 1, year minus 2, and year minus 3 are 0.5, 0.3 and 0.2, respectively.

Relying on FIFA rankings would require a number of adjustments to make the rankings consistent from year to year if we wanted to include all years from 1993 in the sample. For example, without any adjustments, a given change in FIFA points (or a FIFA ranking) from 1993 to 1994 would reflect a different change in performance than would the same change in FIFA points measured from 2008 to 2009. With so many changes to how FIFA rankings and points were calculated over this period, a large number of adjustments would be required. While we will present results of performance models using (unadjusted) FIFA rankings as the dependent performance variable below, we focus on a second national soccer team performance variable – ELO points and rankings – which is not subject to these comparability issues.

ELO points and rankings, named for Arpad Elo who devised this scoring system to rank chess players, assign points to national soccer teams in a given year based on the results of the games during that year. Each national team has a certain amount of points at the end of one year, and the points at the end of the next is just the points from the end of the previous year plus (minus) the points earned (lost) over the year. Year-to-year differences therefore show the total points earned or lost during the current year. Points are earned or lost for each game played depending on: the goal differential, the importance of the game (e.g., an international friendly versus a World Cup qualifying game), and the difference between the actual outcome and the expected outcome. An actual outcome is either 1 for a win, .5 for a draw or 0 for a loss; the expected outcome is determined by strength of opponent and is measured by a continuous value between 0 and 1 based on the pre-game ELO ratings of the two teams. The winning team will get a positive number of points equal to the loss in points earned by the losing team. Under this scoring system, when a team with a high rank plays a team with a low rank, the high (low) ranked team cannot earn (lose) many points if the highly ranked team wins, but the low (high) ranked team will earn (lose) considerable points for an upset. The ELO ranking is simply the ordinal ranking of national teams according to their ELO rating points. While we will present analyses using both FIFA rankings and the ELO measures, a majority of our results will use the ELO measures since they are consistent through the period, they award points depending on strength of schedule, and they are based only on year to year changes. Further, we will focus on ELO ratings rather than rankings since the point scores give more information.

National Team Soccer Players and their Professional Club Affiliations

Central to the analysis of learning and skill acquisition from teammates is a measure of whether national team members are on elite clubs comprised of the best soccer players. The identities of countries' national team players are reported in www.national-football-teams.com. A country-year observation is included in the data set if the www.national-football-teams.com source lists fifteen or more players for a given year The full sample for the analysis contains 101 countries, with a minimum, average, and maximum number of years on a given country's national team of 13, 18.5, and 21 years respectively.

Defining Elite Soccer Clubs

The www.national-football-teams.com source also reports the names of the professional club teams for soccer players on their national teams. This information is used to create various measures of the percent of a national team that play for an "elite" soccer club. Following studies of soccer teams that identify a set of "Big 5" elite soccer leagues, the primary definition used for an elite club in this study will be those clubs that play in these "Big 5" leagues – the English Premier League, Italy's Serie A, Spain's La Liga, Germany's Bundesliga, or France's Ligue Un (Frick, 2009, p. 90). [4]

The empirical analysis also considers additional definitions of elite clubs. A second definition is based on a soccer club's participation in the Champions League. The Champions League is an annual league tournament of top clubs from leagues from around Europe. However, the number of teams participating in the beginning of the league and in the later "group stage" have both varied considerably over the twenty-one year sample period. Using only the Champions League participants or the Champions League group stage participants to identify

[44] There is clear variation in quality across teams in these leagues, but the average quality is higher than in other leagues, and even the teams at the lowest quality level within the league will play the teams at the highest quality level during the season. Further over time a team's quality level will vary and if a team's quality drops far enough given its poor win loss record, it will be relegated to a lower division as high performing lower division teams replace it in the premiere level league.

elite clubs would produce a definition that was systematically more selective in the early years of the sample.

We therefore develop a "Champions Top 50" variable as an alternate definition of elite clubs that keeps the number of elite clubs constant at 50 over the sample period. Under this definition, any club that qualified in the Champions League for the group stage qualifies as elite. Then, to add more teams to the set of elite teams in a year and to keep that number constant, we add the top finishers from the Europa League (the second level club tournament) in the same year, starting with the Europa League finalists, semi-finalists, quarterfinalists, and so on until we have fifty "elite" teams total.[5] This definition of elite clubs therefore changes from year to year depending on which teams are the top finishers in their league.[6] In the empirical work below, we focus on a definition of elite clubs according to the year in which teams participated in the Champions League. While this list is in some later years a more restrictive definition than the list of all Champions League participants but less restrictive in some earlier years, we will refer below to this definition of elite as "Champions League teams." In the empirical analyses, we interact the "Big 5" definition of elite clubs with the "Champions League teams" definition of elite, and identify Big 5 teams that are and are not Champions League teams as well as Champions League teams that are not in a Big 5 league. We consider a "Champions Top 50" team to be elite for the year they are competing in these tournament leagues.

Percent of a National Team on an Elite Team

The independent variable of central interest in performance models below is PCTELITE – the percent of a national team's roster who are members of elite clubs, where elite is measured using the various Big 5 and Champions League definitions. Most analyses below define the PCTELITE variable as the percent of players listed as part of a national's team roster during a given year that are members of elite clubs. If there are n players who play on their national team's games in year t, then each player is 1/n of that roster, regardless of whether the player was

[55] Many Europa league teams are teams that were eliminated from the early play-in rounds of the Champions League so did not qualify for the Champions League group stage. When we have multiple Europa teams to choose from to fill in the last slots of this "top 50" group, e.g. we need to select a small number of teams from among all the Europa League teams who lost in the first round of play, we select the needed number of teams randomly.

[6] There is also some change from year to year in defining which club teams are elite when defining elite teams as those clubs in one of the Big 5 leagues. Each season, two bottom teams are relegated to the second-tier league in the country and replaced by the top performing teams in the second league. The elite status of relegated or promoted teams changes by year.

in all their country's national team's games or in only one game. PCTELITE is just the simple percentage of the roster that is on an elite club team.

For more recent years, this data source reports game by game rosters, so we also conduct complementary analyses where we calculate how many "player-games" there were in a given year (i.e., number of players who played in each game times the number of games). We also calculate the PCTELITE variable as the percentage of these player-games that were manned by players from elite clubs over the course of a calendar year. Again, different definitions for which club teams are and are not elite are used for both the PCTELITE variable based on percentage of the year-long roster and the analogous variable based on player-games.

Timing of Club Participation and National Team Performance

The empirical models below focus in particular on the effects of changes in a national team's PCTELITE variable on changes in ELO points. Therefore the timing of changes in club participation relative to the timing of wins and losses in national team games is important. We focus on models that consider the relationship between (a) changes in ELO points from January 1 to December 31 over year t and (b) changes in team-level PCTELITE variable measured between its value for year t and year t-1. This change in PCTELITE may reflect changes on club rosters that occur within year t itself, and so we are concerned with measurement error in the (change in) PCTELITE variable in these models.[7] For example, consider the case of a player joining an elite club in July as training for a new club season begins. This player will be receiving the "treatment" of being on the elite club which could impact performance in subsequent national team games. However, if that player contributed in important ways to an improved win-loss record for his national team in games from January through June, then the effect of that player on ELO points will not be due to learning but to selection of better performing players since the change in ELO points in this hypothetical example occurs before the player joined an elite club. We therefore re-estimate all models in the empirical section using an alternative specification of the dependent variable that measures the change in ELO points

[7] The www.national-football-teams.com source reports only one club team, the last one, for each national team player in a year, rather than richer data on two (or more) club teams in a year for players who switch club teams.

from July 1 of year t to June 30 of year t+1.[8] Table 1 presents descriptive statistics on key variables from this panel data set on national soccer teams for the 1990-2010 period. All ELO and FIFA measures are calculated using the January 1 dates but the descriptive statistics for the July 1 measures are nearly identical.

Country-Level Time-Varying Covariates

For models that examine the effects of the percent of a national's team players who are members of elite clubs on the national team's performance, we also include several country-specific time-varying covariates. These variables are: GDP per capita, life expectancy, average age of national team players, and a dummy for whether there is a professional soccer club in the country.[9]

Total Number of Foreign Players in Top European Leagues

Empirical models of the effects of elite team membership on national team performance track club membership of national team players. Still, part of the strategy for identifying the effects of membership on an elite club team on the performance of national team players is to rely on the 1995 European Court of Justice's decision (referred to as the Bosman ruling) that reduced dramatically the barriers to foreign-born players joining elite soccer clubs in western Europe. We therefore are interested in whether this ruling caused a discontinuous jump in the number of foreign players on elite clubs.

Part of the analysis to identify any discontinuities in the number of foreign players in these leagues estimates determinants of the PCTELITE variable using the panel data set on national teams. This analysis will address the question of whether the percentages of countries' national teams who are members of an elite European club increase after the Bosman ruling.

[8] Another approach would be to measure the lagged values of all players' club affiliations and use lagged values of the change in PCTELITE calculated from these measures in the models that use January 1 to December 31 changes in ELO points as the dependent variable. However, we cannot measure lagged values of PCTELITE this way because not all of the players on a national team roster in calendar year t are on the roster in calendar year t-1 (ie, PCTELITE in years t and t-1 are by design calculated for different sets of players), and the source of information about club affiliations comes from the source on national football team rosters.

[9] GDP per capita in current US dollars and life expectancy were collected from the World Bank Indicators in July of 2012. The average age of the players was collected from the national-football—teams website. Google searches of national soccer federations produced data on the presence and timing of professional soccer leagues in each country.

However, we supplement this analysis with an examination of data for the top leagues themselves to test for discontinuous jumps in foreign players in top European leagues. While the

Table 1: Descriptive Statistics, National Team Data Set, 1990-2010: Select Variables

VARIABLE	observations	Mean	Std.Dev.	Minimum	Maximum
ELO Rating	1896	1602.6	217.7	883	2151
ELO Rating: Annual Change	1766	.952	48.9	-189	222
ELO Rank	1896	58.50	44.39	1	199
ELO Rank: Annual Change	1766	.133	9.62	-39	40
FIFA Rank	1722	57.46	40.48	1	203
FIFA Rank: Annual Change	1605	-.123	14.37	-67	52
PCTELITE (roster)	1911	.188	.252	0	1
PCTELITE (roster): Annual Change	1783	.133	.067	-.276	.619
PCTELITE (player-games)	1655	.266	.290	0	1
PCTELITE (player-games): Annual Change	1486	.007	.098	-.385	.436
Average Team Age	1874	25.9	1.28	20.3	29.7

data from www.national-footbal-teams.com identify foreign born players who are on their countries' national teams who are playing in top European leagues, it does not identify all foreign players on clubs in the top leagues since it will not cover players who are not members of their countries' national teams. For the league-specific analyses of the effects of the Bosman ruling on the number of foreign born players in these elite European leagues, we need to identify all foreign born players in these leagues including individuals who are not on their national teams. Wikipedia entries for each of the five elite European leagues reports all foreign-born players who ever played in these leagues by year and we use these data to calculate the percent of the EPL, La Liga, the Bundesliga, Serie A, and Ligue Un who are foreign.[10]

IV. National Team-Level Performance Models

To examine whether elite club teams and the peer teammates on those teams can increase the skill and performance of other players, the empirical models relate performance measures of national soccer teams and national teams' players to an "elite club treatment." The models test for improvements in team or player performance on national teams as national team players move to and from elite club soccer teams. The models are designed to rule out other interpretations for any estimated changes in team- or player-performance besides some form of increased learning on top teams.

In this section, we consider the models where national team performance is a function of changes in the number of players on a national team that play on elite clubs. Because the individual player is a member of two teams, we can observe performance on one team (the national team) after we observe changes in membership on his other (club) team. This design feature allows us to track performance for a given team (his national team) while the quality of teammates he is exposed to during club play varies from year to year. Using widely agreed upon definitions of which teams are elite (e.g., Champions League clubs or English Premier League clubs), we measure the timing of club membership and changes in club membership carefully so that changes in the PCTELITE variable measure changes in the "club team treatments" during the year for which national team performance is measured.

[10] The data was collected in September 2011 from the following Wikipedia pages:
http://en.wikipedia.org/wiki/List_of_foreign_Premier_League_players;
http://en.wikipedia.org/wiki/List_of_foreign_La_Liga_players;
http://en.wikipedia.org/wiki/List_of_foreign_Bundesliga_players;
http://en.wikipedia.org/wiki/List_of_foreign_Ligue_1_players;
http://en.wikipedia.org/wiki/List_of_foreign_Serie_A_players

Our initial models express team performance as a function of the percentage of national team players who played on elite clubs during that year while controlling for country (national team) fixed effects:

$$\text{ELO-POINTS}_{ct} = a + b\ \text{PCTELITE}_{ct} + c\ \mathbf{X}_{ct} + \mu_c + \varepsilon_{ct} \qquad (1)$$

ELO POINTS_{ct} are the total ELO ranking points for the national team from country c by the end of year t. As discussed in Section III, the construction of the dependent performance variable, ELO POINTS, already adjusts for a number of factors including the quality of the opposing teams played by country c in year t, the importance of the soccer matches played, and other factors. PCTELITE_{ct} is the total number of players on country c's national team roster that played on an elite team (initially defined as any club in the top division of the English, Spanish, Italian, German or French leagues) in year t divided by the total number of players ever rostered on the team in year t, **X** is a vector of time varying covariates, μ is a country fixed effect, and ε is the error term. With the country fixed-effects included, b in equation (1) therefore measures the effect of within-team changes of PCTELITE on within-team changes in ELOPOINTS, and provides an initial test of whether a given national team's ELO improvement (decline) is bigger in years when more (less) of its roster is a member of an elite club.

When variants of this fixed effects model are estimated, the results in Table 2 are obtained. In the column 1 sample based on up to 21 years of data from 101 countries' national teams, the estimated effect of PCTELITE on ELO POINTS is significant and positive, indicating that in years when a given national team has a larger share of players playing on elite European club teams, the national team performance improves. The estimated coefficient of 178.3 implies that a one standard deviation increase in PCTELITE (25 percentage points[11]) will increase ELO POINTS earned in the following year by 44.6 points, or roughly the difference between the 35th ranked team and the 46th or 26th ranked team in 2000.

Still, in these models with country fixed effects, the estimated effect of PCTELITE may simply reflect superior national team performance in a set of years when a larger number of rostered players were on elite European club teams and poorer performance in some other period

[11] For simplicity we use the measures of standard deviation presented in Table 1. While samples change in each analysis, we find that the means and standard deviations of these variables remain quite similar.

several years removed from the first period when few players were on the top clubs. Or put differently, country fixed effects do not ensure roster fixed effects given the long twenty-one year time span considered in these models. One would prefer to watch a team of the same players from year t to year t+1 and match ELO changes in those years to PCTELITE changes in those same years. As a first step, we restrict the sample of years considered in columns (2) – (6) to include only those years where the roster in year t+1 (the following year) had some threshold percentage of common players on the national team roster, ranging from 40% of the players in common in column (2) to 80% of the players in common in column (6). In each equation, the effect of PCTELITE remains positive and significant. In the column (6) model where the sample size is reduced to 307 national team-year observations, the coefficient on PCT ELITE of 205.4 implies that a one standard deviation increase in PCT ELITE would lead to an increase of 51.4 ELO POINTS.

Table 2: Effect of Elite Players on National Team ELO Ratings as Roster Matching Increases

(All models include country fixed effects and time varying controls for GDP per capita, life expectancy, a dummy for whether the country has a professional soccer league, average age of the team and average age squared. Standard errors are clustered by country.)

VARIABLES	(1) Elo Rating Full Sample	(2) Elo Rating % same >=0.4	(3) Elo Rating % same >=0.5	(4) Elo Rating % same >=0.6	(5) Elo Rating % same >=0.7	(6) Elo Rating % same >=0.8
% Roster Elite	178.3***	117.8***	122.7***	135.4***	160.5***	205.4***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1,896	1,711	1,555	1,218	721	307
R-squared	0.138	0.117	0.121	0.139	0.160	0.265
# of countries	101	101	101	101	101	87

Robust p values in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Because of the important time component of the effect we are trying to measure, we use a first differencing of equation 1 to provide a cleaner comparison of year-to-year changes in PCTELITE to the corresponding change in ELO performance.

$$\Delta \text{ELO-POINTS}_{c(t,t+1)} = a + b\ \Delta\text{PCTELITE}_{c(t,t+1)} + c\ \text{ELO-POINTS}_{t} + d\Delta\mathbf{X}_{c(t,t+1)} + \varepsilon_{c(t,t+1)} \qquad (2)$$

As the players on a given national team switch from not being a member of an elite club to being a member of an elite club, or vice versa, we can measure whether the national team playing with many of the same players from one year to the next increases or decreases its ELO point total. This specification eliminates any possibility that the coefficient on the PCTELITE variable reflects differences in performance of very different teams playing many years apart. The equation (2) models also include a control for base year ELOPOINTS since ELO points by design cannot increase (decrease) much for high (low) ranked teams.[12] As in Table 2, we also estimate the equation (2) specification with increasingly restrictive samples where the number of players in common in sequential years increases.

Table 3: Effect of Change in Elite Players on Change in National Team Elo Ratings

Full Sample and as Roster Matching Increases

(All models have robust standard errors and include a lagged value of Elo rating and controls for change in GDP per capita, change in life expectancy, change in whether the country has a pro soccer league, and change in average age and average age squared of the team.)

VARIABLES	(1) Δ In Elo Full Sample	(2) Δ In Elo % same >=0.4	(3) Δ In Elo % same >=0.5	(4) Δ In Elo % same >=0.6	(5) Δ In Elo % same >=0.7	(6) Δ In Elo % same >=0.8
Δ in Elite	66.17***	36.79**	41.67***	45.32**	39.73*	107.86***
	(0.000)	(0.014)	(0.007)	(0.011)	(0.086)	(0.003)
Observations	1766	1598	1454	1139	669	278
R-squared	0.021	0.016	0.017	0.016	0.014	0.045

Robust p values in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimated coefficients on the PCTELITE variable in the equation 2 specification are reported in Table 3. The PCTELITE coefficients are always positive and significant. While the evidence across all specifications is very consistent, the model in Table 3 column 6 model reduces the time between observations and keeps roster membership very constant,[13] thus

[12] All models in Table 2 were estimated without the base year ELO points variable included and all PCTELITE coefficients remain positive, significant, and similar in magnitude to the corresponding model in Table 2. We report the models with lagged ELO points included given the way ELO points are awarded to high and low ranked teams.

[13] Since it is likely that national team's may be changing players at the margin of their roster, having less than a 20 percent difference in the identities of players who played in some of their national team's games is a small

providing strong evidence that *players who join or leave elite clubs are responsible for the subsequent changes in their national team's performance*. In the column 6 specification, the PCTELITE coefficient implies that a one standard deviation increase in the change in percentage of a national team's roster that becomes a member of an elite club over a given year would change the national team's ELO points by 7 points in the following year. In most places in the ELO POINTS distribution, 7 points will lower the team at least one ranking, and in some places in the distribution it will lower the team as many as 4 rankings spots.

To be clear, the results in Table 2 and especially Table 3 are not simply indicating that teams with better players – those on top club teams – have better performance. That is undeniably true.[14] Rather, because models reported in Table 3 are examining changes in team performance in contiguous years, and because models also consider cases where rosters do not change, the models above are looking at what happens to the national team when a given player switches to or from an elite club. Even when rosters are essentially the same from year to year, when a given national team player joins (leaves) an elite European club, team performance improves (declines) the next year.

Models with Alternative Measures of Team Performance and Elite Club Team Membership

We now consider whether the results in Table 3 are robust with respect to different measures of team performance and alternative definitions of what club teams are elite clubs. In the next set of models, we first consider an alternate measure of PCTELITE based on the number of player-games across all national team games in a year played by elite and non-elite club players rather than the number of players ever appearing on a roster in a given year. We then consider alternate definitions of which clubs are and are not elite, including our Champions League ("Top 50") definition of which clubs are elite in a given year. Finally, we measure national team performance using the more commonly discussed FIFA rankings as well as ELO rankings in place of ELO points.

---

percentage that may often involve players who receive little playing time. A precise examination of this is beyond the scope of the roster data.

[14] In simple cross section models of elo rating on percent roster elite, the coefficient is 542.04 with a t statistic of 43.12. Teams whose percent elite is one standard deviation above the mean have elo ratings 146 points above the mean which is comparable to a 31 point jump in a team's rank in 2000.

Table 4 column 1 replicates the full panel sample model from Table 3 column 1 for comparison in which: the dependent variable is change in ELO POINTS; PCTELITE is the change in the percentage of all the players who were ever rostered on the national team in a given year who played on a club in one of the top five leagues (row 1); and the sample is the full set of national-team years from 1990-2010. In the columns 2, 4 and 6 specifications, we measure PCTELITE in a given year by the total number of player-games during a national team's year that is played by elite club players (row 2) . For example, if a national team plays ten games in a given year with no substitutions in any game, there are 110 player-games. If two stars who played on elite clubs played in all ten games, another star player who is an elite club member played in five games, and all other positions in all ten games were accounted for by players who were not on elite clubs, PCTELITE under this alternate player-game definition would be 22.7% (i.e., 25/110). If there was some swapping in and out of the roster from game to game among the non-elite club members on this national team so that there were 50 different players rostered over the course of that year, the PCTELITE variable would have been only 6% (i.e., 3/50) under the previous definition. In short, this revised definition of PCTELITE will consider a team as having more elite players than the prior definition based on roster spots if elite players play more games and there is some churning among the non-elite club players who are rostered. This player-game based definition of PCTELITE can only be calculated in later years when game by game data on rosters is available, and so is available for a smaller sample of country-years than the former definition. The definition of an elite club in Table 3 remains as before – any club in the English, Spanish, Italian, German, or French top leagues.

When this alternative definition of PCTELITE based on the percent of player-games, rather than roster spots, that are taken by players on elite clubs is used to calculate how much change there is in the extent of elite club representation on a country's national team, the results in column 2 of Table 4 are obtained. The coefficient on the change in PCTELITE variable remains positive and significant, with the coefficient implying that a one standard deviation increase in this measure of PCTELITE leads to a 4.6 point increase in ELO rating over a given year.

Table 4: Effect of Change in Elite Players on Change in Elo/Fifa Ratings and Rankings Alternative Specifications

(All models have robust standard errors and include a lagged value of the dependent variable and controls for change in GDP per capita, change in life expectancy, change in whether the country has a pro soccer league, and change in average age and average age squared of the team.)

VARIABLES	(1) Δ in Elo rating	(2) Δ in Elo rating	(3) Δ in Elo rank	(4) Δ in Elo rank	(5) Δ in FIFA rank	(6) Δ in FIFA rank
Δ Elite-Roster	66.18***		-9.711***		-11.70**	
	(0.000)		(0.001)		(0.050)	
Δ Elite-Player Games		46.67***		-5.454**		-12.14***
		(0.001)		(0.021)		(0.001)
Observations	1,766	1,475	1,766	1,475	1,605	1,371
R-squared	0.021	0.023	0.015	0.014	0.039	0.053

Robust p values in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the remaining four columns of Table 4, we examine the effects of PCTELITE measured in terms of percent of the roster and percent of player-games accounted for by elite players but with different dependent variables to measure national team performance. In the columns 3 and 4 specifications, we use the annual change in ELO rankings (the ordinal ranking of teams by their ELO points), while in columns 5 and 6 the dependent variable is the annual change in FIFA rankings. As discussed in section III, FIFA rankings differ from ELO rankings because they are weighted averages of several years of prior rankings, the ranking system has changed several times over the period, and the rankings are not available prior to 1993. For both ELO and FIFA rankings, the best teams with the most points are the teams with the lowest numerical rankings. All results in Table 4 show consistently that increases in the extent to which a national soccer team has players who are members of an elite club immediately increases the subsequent performance of the national team, regardless of which of the two different PCT ELITE measures or three different team performance measures are used in the models.

In Table 5, we consider alternate definitions of which professional clubs are "elite." We expand the definition of "elite" clubs and consider our "Champions League" definition for identifying elite club teams. We interact this definition with our prior definition of any club in the top five European leagues and now use the following three dummy variables to define elite clubs: (a) a club team which is in one of the top five European leagues and which is competing in the Champions League; (b) a club team which is competing in the Champions League but which is not in one of the five top club leagues; and (c) a club team which is in one of the top five European leagues but which did not qualify for the Champions League. Variables (a) and (c) together combine to form the category of any club in one of the top five leagues used to define elite in the prior three tables, while players who were members of the teams identified by variable (b) were considered to be members of non-elite teams in the Table 2-4 models. Because Champions League teams change from year to year, the model will be examining whether specific pairs of years of a national team's performance show especially high (low) changes because more (less) of the team participated on Champions League teams.

The three columns of Table 5 again use three different dependent team performance variables: ELO points, ELO rank, and FIFA rank. The models consider changes in performance as a function of changes in the percentages of the national team who in the season beginning in the prior fall were members of the three different categories of potentially elite teams. The results in Table 5 show a consistent pattern. The point estimate of the coefficient on the percent of a national team who were members of a Champions League team from one of the top five leagues is the largest in magnitude in two of the three models. The effect of being on a club team in one of the top five leagues but not a Champions league team also remains significant. F-tests reveal that in all three models, these two coefficients – for being on clubs from a "Big 5" league who are and are not Champions League teams – are not significantly different from each other at conventional significance levels. However they are significantly higher than the coefficients on the variable measuring membership on Champions League teams that are not in the "Big 5" leagues. In the ELO rating and ranking models, there is still some positive but very small effect on national team performance of changing memberships of players on the best teams outside of the Big 5 Leagues.

These results suggest that some of the benefit from club membership is the caliber of the competition the club team plays. Even the bottom teams in La Liga or the EPL play Barcelona and Real Madrid, or Manchester United and Chelsea, each year. Moreover, the results lend some additional support for the way in which the previous tables defined elite clubs – any team in the top five European league and not just the top finishers. New membership on one of the top finishers in a Big 5 league appears to have a similar effect on national team performance to the effect one observes for new membership on other teams in the Big 5 leagues according to the results in Table 5. These effects are bigger and more significant than the effects of new membership on Champions League teams from lesser leagues.

Table 5: Effect of Change in Elite Players (Defined According to Champions League Variables) on Change in Elo Rating, Elo Ranking and Fifa Ranking

(All models have robust standard errors and include a lagged value of the dependent variable and controls for change in GDP per capita, change in life expectancy, change in whether the country has a pro soccer league, and change in average age and average age squared of the team.)

VARIABLES	(1) Δ in Elo rating	(2) Δ in Elo rank	(3) Δ in Fifa rank
Δ in roster players in top 5	80.51***	-10.06***	-14.47**
leagues and in champions league	(0.001)	(0.003)	(0.018)
Δ in roster players in top 5	66.30***	-10.32***	-10.39*
leagues but not in champions league	(0.000)	(0.001)	(0.094)
Δ in roster players not in top 5	25.07**	-3.940*	3.184
leagues but in champions league	(0.034)	(0.100)	(0.422)
Observations	1,766	1,766	1,605
R-squared	0.023	0.017	0.040

Robust p values in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Measuring the Contribution of Elite Club Players Just Before and After They Join Their Elite Club

The Table 3 models, especially a specification like the one in Table 3 column 6 that keeps the composition of national team rosters consistent from one year to the next, provide evidence of a significant treatment of the treated effect that players receive after joining an elite team. Models in Tables 4 and 5 demonstrate that the finding of a treatment of the treated effect associated with joining an elite soccer club is robust with respect to alternative measures of both the dependent team performance measure and the key independent variable measuring membership on elite clubs.

These models incorporate several features that address the issue of selectivity of stronger players being chosen by elite clubs. The dependent performance variable measures the change in national team performance between year t-1 and year t. If a player who joins an elite club in year t was contributing more to his national team's results in year t-1 than other players did, this would be reflected in a higher year t-1 ELO ranking. This again underscores one of the useful features of the setting of professional soccer for identifying the productivity effects due to teammates. Better players on the national team may be the ones that top soccer clubs add to their rosters, but these players are still members of their national team after they join the elite club. In this section, we report results of additional models that take account of this type of selectivity in an even more explicit way. In particular, we measure the contribution of elite club team members to their national teams in the year just before they join an elite club and compare it to the contribution of exactly the same players in the year they join their elite club. That is, we compare the contribution of "about to be first time elite players" in year t to improvement in the contribution of exactly the same individuals when they are "first time elite players" in year t+1.

To make this comparison, we first identify four different categories of national team players. In year t, we identify players who were on their national team in both year t and year t-1, who were on an elite club team in year t (i.e., in the fall prior to year t), but who were never on an elite club team before year t. In year t, these are "first time elite club players" who were also contributing to their national teams in year t-1. They are identified as group A. The coefficient on the variable measuring membership in group A in the equation below is $\beta_{1sttimerveteran}$. A second group of players are those in year t who play for their national team and are not on an elite team but will be on an elite team next year. These are "about to be first time elite players." These players make up group B in the equation below with coefficient $\beta_{Nextyear}$. A third group is comprised of those players who are veterans to their national teams and veterans to elite clubs in year t. We label this group C. A fourth group is comprised of players who are new elite players in year t but also new to the national team. These players whom we label group D are different from the group A players in that the group D players did not contribute to the performance of their national team in year t-1. If we rewrite the change-in-ELO-points equation (2) above to incorporate these four groups of national team players, we obtain:

$$\Delta ELO_{t,i} = \beta_0 + \beta_1 ELO_{t-1,i} + \beta_2 \Delta X_{t,j} + \beta_{Nextyear} B_{t,i} + \beta_{newelite} D_{t,i} + \beta_{1sttimerveteran} A_{t,i} + \beta_{veteranelite} C_{t,i} + e_{t,i} \quad (3)$$

In equation (3), we concentrate on the coefficients $\beta_{1sttimerveteran}$ and $\beta_{Nextyear}$. If elite club players are already showing elite level talent in the year before they are selected for an elite club, then the coefficients on both variables will be positive and significant and of comparable magnitude.[15]

Results obtained when equation (3) is estimated are shown in Table 6. The coefficient on first time elite players (row 2) is one and two thirds times the magnitude of the coefficient on next year elite players (row 1). An F-test rejects the hypothesis that these two coefficients are equal at less than the 0.01 level.

Table 6: Effect of First Time Elite Players and Next Year Elite Players on Change in National Team Elo Ranking, 1990-2010

(The model includes a lagged value of elo rating, controls for new elite players and veteran elite players, time varying covariates and robust standard errors.)

VARIABLES	(1) Δ in Elo Rating
% of veteran or first time Roster players who will be first time elite next year	67.14*** (0.001)
% of veteran Roster players who are first time elite this year	112.59*** (0.004)
Observations	1,707
R-squared	0.037

Robust p values in parentheses; *** p<0.01, ** p<0.05, * p<0.1

This pattern of results supports two conclusions. First, when elite clubs add players who have no previous elite club experience to their team's roster, they select, as expected, individuals who are better performers than other players on the team. Second, and most important for the focus on this paper, these same players contribute even more to their national teams in the year after they are members of an elite club team than in the preceding year, despite the fact that they were relatively high performers prior to joining the elite club.

One potential concern with the equation (3) specification reported in Table 6 is that the relationship between the change in ELO dependent variable and the variables measuring the

[15] The previous results show in Table 3 column 6 results suggest that we will find that the "first timers" have a bigger effect on ELO than do "about to be first timers" since, in models with the roster staying close to constant, an increase in percent elite results in a higher change in ELO.

percent of the roster that is first time elite players and soon to be elite players is in part determined by cross-team variation. In particular, equation (3) does not explicitly link the "about to be first time" elite players' impact on the change in ELO points in year t-1 to the change in ELO points in year t for the exact same national team. We really want to see how a player on one team changes that team's performance when he becomes elite for the first time.

To make this precise comparison, consider again the equation (3) specification and how it will change by period t+1 after one more year of changes occur in a national team's roster. Group B from period t (the group of about to be first time elite club players) has become veteran national team players and are now first time elite club players. Groups D and A from period t (where A was the group of players in year t who were first time elite club members and who were also on the national team the prior year t-1; and D was the group of players in year t who were first time elite club members but who were not on the national team the prior year) have now become veteran elite in year t+1 (i.e., veteran players on both the elite and the national teams). There are new groups to define with the additional year of changes in year t+1 – group E has the national team players who will be elite in year t+2, and group F has new national team players who are already elite. Thus, we can write the change in ELO model for period t+1 as:

$$\Delta ELO_{t+1,i} = \beta_0 + \beta_1 ELO_{t,i} + \beta_{Nextyear} E_{t+1,i} + \beta_{newelite} F_{t+1,i} + \beta_{1sttimervetean} B_{t+1,i} + \beta_{veteranelite}(A_{t+1,i} + C_{t+1,i} + D_{t,i}) + e_{t+1,i} \qquad (4)$$

Subtracting equation (3) from equation (4), one obtains:

$$\begin{aligned} &\Delta ELO_{t+1,i} - \Delta ELO_{t,i} = \beta_1(ELO_{t,i} - ELO_{t-1,i}) + \beta_{Nextyear} E_{t+1,i} + \beta_{newelite} F_{t+1,i} + \\ &(\beta_{1sttimer} - \beta_{Nextyear}) B_{t+1,i} + \beta_{veteranelite}(A_{t+1,i} + D_{t+1,i}) + e_{t+1,i} \end{aligned} \qquad (5)$$

Equation (5) allows us to estimate directly the difference between the effects on a given national team's performance of the same player(s) from just before to just after he (they) joins an elite team. When the dependent variable is the change in the change in ELO in equation (5), the coefficient on the variable $B_{t+1,i}$ that measures the percent of players who are first time elite club members (but veterans on their national team) in year t+1 yields the parameter of interest – the difference between the effect of the percent who are veteran first timers and the effect of the percent who are first timers next year, $\beta_{1sttimer} - \beta_{Nestyear}$.

When we estimate equation 5, this coefficient is positive and significant with a point estimate of 66.7. The equation (5) estimate of the coefficient on the variable for membership in group $E_{t+1,i}$ measures the effect of these exact same players but in the year just before they become elite players. This is also positive and significant with a point estimate of 51.0. Since the former coefficient on the variable for veteran first timers measures the difference between the coefficient on veteran first timers and next year first timers in a given year, these two estimates imply that the contribution to the national team of "veteran national team members who are first time elite club members" is more than twice the contribution that these same players made in the year just before they joined an elite club.

We can never test whether elite club teams somehow find players who will show this acceleration in performance all on their own in their next year after they join an elite club. The results here do however test whether the contribution of national team players who become members of an elite soccer club is no different from the year just before they join the elite club to the year after they join the elite club. The evidence runs counter to this pattern. Scouts for elite clubs are not reacting to any such observable evidence in their selection decisions. Scouts for elite club teams do identify players who are contributing to their national team's performance at levels that exceed the contribution of other national team players who do not play on elite clubs. However, the new players they select for their clubs contribute at significantly higher levels to their national teams after they have benefited from an "elite club treatment."

V. The Bosman Ruling and Its Effects on Growth in Foreign Players on Elite Clubs

The results in Tables 2-6 tell a consistent story. When the number of players on a national team who become members of an elite club in one of the top European leagues increases, the performance of that national team improves in the next year. The estimates from these models, especially models like those in Table 3 column 6 that keep rosters as fixed as possible, or the estimates in Table 6 which examine the contribution of exactly the same players just before and just after they join an elite club and examine changes in team performance from one year to the next, minimize effects due to omitted variables. Still, these carefully constructed estimates of the effects of membership on an elite club team are estimates of treatment-of-the-treated (players) effects.

In this section, we estimate models that take advantage of a unique period in professional

European soccer characterized by an exogenous increase in the number of foreign players who became members of elite European club teams to develop estimates of local average treatment effects of the PCTELITE variable. In particular, we will focus on the period before and after the 1995 ruling of the European Court of Justice in the Bosman case. This landmark ruling removed restrictions on the (previously low) number of foreign-born players who could be members of professional soccer clubs in Europe. If this ruling was responsible for a discontinuous jump in the probability of the key "treatment" in this study – becoming a member of an elite club – then the switch from the pre- to post-Bosman periods can serve as an instrument in a fuzzy regression discontinuity design. These models address a different aspect of selectivity concerns than the previous models do by measuring the effect of the PCTELITE variable on a set of players who likely come from a lower part of the talent distribution of players.

The Bosman Ruling

The Bosman ruling by the European Court of Justice in 1995 adjudicated and settled the case of Jean-Marc Bosman who was prevented from joining a French soccer club because his Belgian club demanded a large transfer fee.[16] The 1995 ruling was a landmark decision, and according to many it was an unexpected reversal of the long-standing ownership rights enjoyed by European soccer clubs. Prior to Bosman, soccer leagues followed a "3+2" rule to determine how many foreign players could be on a team. The limit was three players, with some exceptions made for two other foreign players if they had been developed by a club team for many years. The Bosman ruling struck down this limit. Moreover, the European Court ruled illegal the long standing practice that allowed players to move between soccer clubs only after the last team to employ the player agreed with some new club on a transfer and the fee for the transfer, stating that both of these restrictions violated labor law provisions in the Treaty of Rome. While these restrictions were eliminated for the fifteen EU countries then covered by the Treaty of Rome, clubs could still employ three (or 3+2) foreign players from outside of the 1995 EU member nations. Some pre-existing player contracts did extend past 1995, but this 1995 ruling eliminated any meaningful restrictions on the inclusion of foreign players in European leagues in EU nations including all of the Big Five Leagues.

---

[16] Kleven, Landais, and Saez. (2013) exploit this change in legal regimes to investigate the effects of tax policies on migration of soccer players.

Testing for Discontinuity in the Number of Foreign Players in Elite Leagues after Bosman

To validate the regression discontinuity design, we first examine whether the Bosman ruling caused a discontinuous jump in the number of foreign players in the top European leagues. We do this in two ways. First, we examine the percent of foreign players in the English, Spanish, Italian, German and French leagues before and after the 1995 Bosman ruling. These models are estimated in league-based samples. Next, because our empirical models focus on the effects of the PCTELITE measure on the performance of national soccer teams, we also test for discontinuities in the percent of national team players who play in the top European leagues.

Figure 1 shows the percentage of foreign players in the Elite 5 leagues by season. Some of these foreign players are members of their home countries' national teams, while others are not.[17] This graph shows that the steepest growth in the number of foreign players in these leagues is precisely at the time of the Bosman ruling. Figure 1 reveals another important fact. Teams in these leagues are near the maximum number of foreign players allowed in the year before the Bosman ruling. In 1995, the percent of players who are foreigners in the Big 5 leagues is approximately 15%. Assuming a roster size of 25 players in 1995, a limit of 3 foreign players would be 12% of all players and a limit of 5 players would be 20%. Assuming an average roster size of 30, these percentages would be 10% and 17% respectively. Teams league wide appear to have bumped right up against the 3 (or 3+2) player limit before the Bosman ruling.

To test formally for discontinuity in the percentage of foreign players in these leagues between 1995 and 1996, we estimate models that express the percentage of players in the top five European leagues who are foreign players as a function of a Bosman ruling dummy variable, which equals one for all years after 1995. Table 7 presents several different models that include different specifications for a 1990-2010 time trend and for a post-Bosman time trend. In all specifications considered, regardless of the specification of the overall and post-Bosman time trends, the estimates show that the Bosman ruling caused a discontinuous jump in the percent of

[17] See section III for data sources that identify foreign players in these leagues who are not national team members.

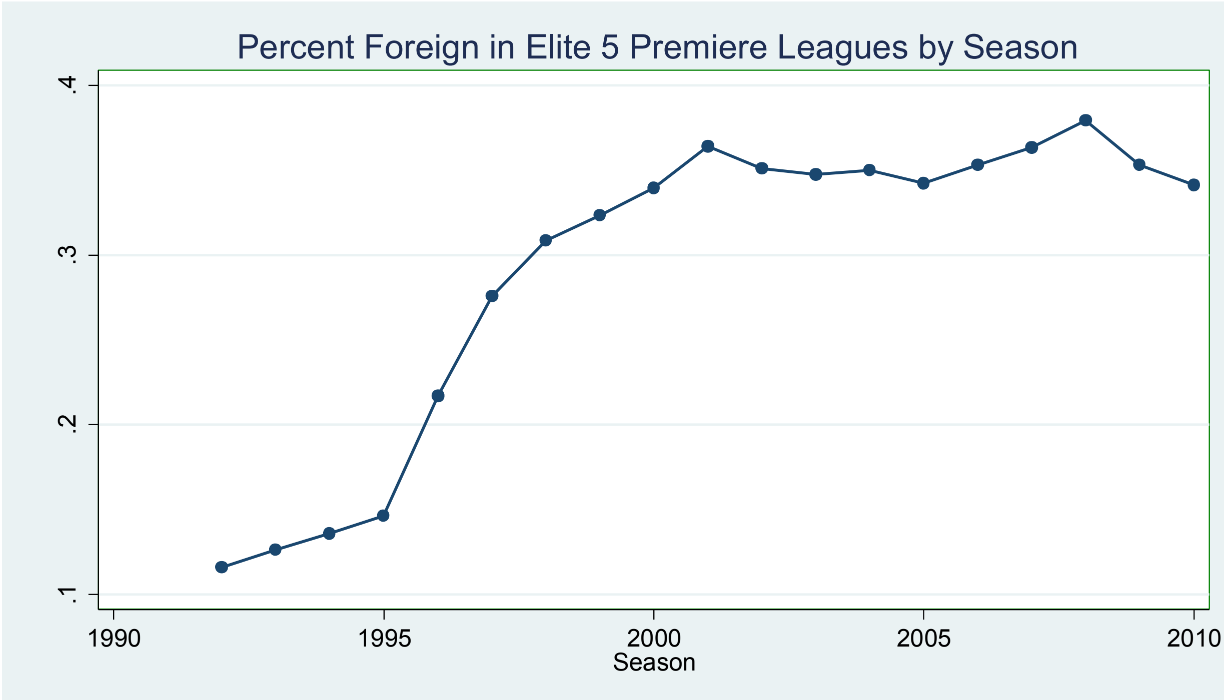

Figure 1: Percent of Foreign Players by Year in the Top Divisions of the Five Top European Leagues

foreign players in these top leagues of 8 to 14 percentage points.[18]

While Table 7 examines whether the Bosman ruling increases the employment of foreign players in the Big 5 soccer leagues, we are also interested in the existence of a discontinuous jump in the percent of the rosters of the 101 national soccer teams that play in these top leagues. The results shown in Table 7 could be due to increasing numbers of foreign players in these top leagues after Bosman from only a small number of soccer strongholds. If so, there would not be a discontinuous jump in the number of national soccer team players who join these leagues in many or even in most countries. We therefore test whether the percentage of national soccer team players who play in these leagues also exhibits a positive discontinuous jump between 1995 and 1996.

[18] Separate regressions for each of the five leagues considered also show large significant positive effects of the Bosman ruling on the number of foreign players in each league. The coefficients on the post Bosman variable in all five models are positive and significant for models that examine the percent of foreign players in the English, Spanish, and Italian leagues. Four of the five coefficients on this variable are positive and significant in models estimated on data for the German and French leagues, and all five coefficients have positive point estimates.

Table 7: Effect of Bosman on % of Foreign Players in Top Leagues
(All models include league fixed effects, and standard errors are clustered by league)

	(1)	(2)	(3)	(4)	(5)
Bosman	**0.131*****	**0.102****	**0.101****	**0.140*****	**0.080****
	(0.007)	**(0.014)**	**(0.018)**	**(0.007)**	**(0.037)**
Time Trend (TT)	0.007	0.017*	-0.009	0.002	0.002
	(0.105)	(0.093)	(0.589)	(0.786)	(0.787)
TT^2		-0.000	0.002		
		(0.215)	(0.230)		
TT^3			-0.000		
			(0.179)		
PostBosman TT				0.005	0.026
				(0.592)	(0.112)
PostBosman TT^2					-0.001*
					(0.056)
N	103	103	103	103	103
adj. R^2	0.725	0.728	0.744	0.724	0.753

Robust p values in parentheses
***p<0.01, **p<0.05, *p<0.1

Table 8 replicates the models but for the sample of national team rosters. In particular, the models in Table 8 report results of models that express the percent of a national team's roster that are members of teams in these top European leagues as a function of various time trends and a post-Bosman (post-1995) dummy. Regardless of the specification of the overall time trend, the post-1995 time trend, and inclusion of country specific time trends, the results consistently show a large permanent discontinuous jump in the number of players who are members of teams in these elite European leagues. The magnitude ranges from 1.8 to 6.3 percentage points. By way of illustration, if rosters equal 20 players, a five percentage point increase would amount to one more player on average in each country's national team who is a member of a club team in one of these elite leagues. Whether we look at the increase in the numbers of foreign players in these top five leagues or at the number of different countries' national teams' players who gain membership in one of these leagues, we observe a clear spike in foreign players in these leagues right after the Bosman ruling was handed down.

Table 8: Effect of Bosman on Elite Players on National Teams
(All models include country fixed effects, and standard errors are clustered by country.)

VARIABLES	(1) % Roster Elite	(2) % Roster Elite	(3) % Roster Elite	(4) % Roster Elite	(5) % Roster Elite	(6) % Roster Elite
Bosman	**0.035******	**0.019****	**0.018***	**0.063*****	**0.032*****	**0.050*****
	(0.000)	**(0.047)**	**(0.055)**	**(0.001)**	**(0.002)**	**(0.009)**
TT	0.003***	0.009***	0.014**	0.009***	0.008***	0.012***
	(0.002)	(0.004)	(0.014)	(0.002)	(0.000)	(0.000)
TT^2		-0.0002**	-0.001			
		(0.038)	(0.193)			
TT^3			1.45e-05			
			(0.368)			
Bosman*TT				-0.007**		-0.004
				(0.033)		(0.168)
Country Specific Time Trends	No	No	No	No	Yes	Yes
Observations	1,911	1,911	1,911	1,911	1,911	1,911
R-squared	0.911	0.911	0.911	0.911	0.947	0.947

Robust p values in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VI. Two-Stage Regression Discontinuity Model Estimates

With consistent evidence that the Bosman ruling caused a discontinuous upward shift in the number of foreign players being selected onto teams in the top five European leagues, we now return to the models of the effects of the PCTELITE variable on national team performance with a special focus on the effects of the PCTELITE variable right around the time of the Bosman ruling when selectivity of foreign born soccer players was significantly and dramatically relaxed. The second stage model has the same form as the models previously shown in Table 3, and the first stage model is estimated using the column 5 specification from Table 8. Because the variables are changes rather than levels, the change in percent elite is instrumented with a dummy for 1996 when Bosman took effect plus country fixed effects.[19] The models and samples in two stage fuzzy regression discontinuity models therefore focus on the performance effects of growth in PCTELITE due to the Bosman ruling.

[19] Note that the column 5, Table 8 specification is $PCTELITE_{it}=\beta_0 +\beta_1 (Bosman)_t+ \beta_2(time)_t+\beta_3(country)_i + \beta_4(country_i*time_t)+ \varepsilon_{it}$. Transforming this equation into year to year changes $\Delta\ PCTELITE_{i,t-(t-1)}= \beta_2 + \beta_1\Delta(Bosman)_{t-(t-1)}+ \beta_3(country_i) *\Delta time_{t-(t-1)}$ where the only non-zero Bosman change is for 1996 and the time trend changes become 1 so country fixed effects remain.

As a simple introduction to these two-stage models, column 1 of Table 9 provides OLS estimates without instruments for effect of the PCTELITE variable, but for precisely the two-year period on either side of the Bosman ruling. Since this model examines the effects of changes in PCTELITE from just before to just after the Bosman ruling, it estimates the effects of growth in PCTELITE when much of the change in PCTELITE is very likely due to the Bosman ruling. Estimates from this simple model should provide a good indication of what the two stage models will show. The results in the Table 9 column 1 model for change in team performance over this two-year period reveal that the effect of PCTELITE is positive and significant.

Table 9: Effect of Change in Elite Players on Change in Elo Rankings:
IV Estimation with Bosman Ruling as Instrument

(Country fixed effects are also used as instruments, and time varying covariates are included. Errors are clustered by country)

VARIABLES	(1) 1995-1996 Δ in ELO Rating	(2) All Years Δ in ELO Rating	(3) 1990-1996 Δ in ELO Rating	(4) 1994-1996 Δ in ELO Rating	(5) 1994-1997 Δ in ELO Rating
Δ in Percent Elite	103.0* (0.100)	1,214*** (0.000)	278.3*** (0.000)	276.5*** (0.004)	264.1*** (0.007)
Instrumental variables	No	Yes	Yes	Yes	Yes
Observations	86	1,766	389	230	327

Robust p values in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In columns 2 through 5 of Table 9, we present estimates of the PCTELITE effects from various two stage models. Column 2 gives the estimates from the instrumental variables model estimated on the full set of years from 1990-2010. In the samples for columns 3 and 4, we include 1996 as the lone post-Bosman year. Column 3 uses the full set of pre-Bosman years in our sample (1990-1995) while column 4 includes only 1994 and 1995 as the pre-Bosman years. These two different samples with the maximum and minimum number of pre-Bosman years respectively will therefore have different estimates of the pre-Bosman time trend in PCTELITE. These models eliminate any need for controls for a post-Bosman time trend by focusing on the effects of Bosman on PCTELITE in1996. Column 5 considers two pre- and two-post Bosman years to extend the number of years when foreign players continued to enter the elite leagues (as shown in Figure 1) while still keeping the window of years around Bosman very narrow.

In all samples and specifications, the estimated effect of changes in PCTELITE on changes in ELO performance is positive and significant. The magnitude of the estimates in columns 3 through 5 that consider very narrow windows of one or two years after Bosman are all very consistent. The estimated coefficients there imply that a one standard deviation increase in change in PCTELITE would increase change in ELO by 18 points, which in turn translates into a change in ranking of as high as 10 spots and as low as 2 depending on where the team is in the rankings. These results in Table 9 provide a stronger case for interpreting the estimated PCTELITE effect as causal.[20] *Effects of exposure to elite clubs continues to have an impact on subsequent performance of the players' national teams even when models focus on the effects of PCTELITE for the expanded set of, presumably somewhat less talented, foreign players who joined elite clubs after the Bosman ruling.*

VII. Potential Measurement Error in the Timing of Elite Club Membership

All of the estimates of treatment-of-the-treated effects reported in Tables 2 through 6 and the estimates of local average treatment effects in Table 9 are based on models that analyze changes in the national teams' PCTELITE values from one year to the next and relate those changes to changes in team performance over the same period. We now consider the possibility that the timing of changes in elite club membership and changes in the win rates of national teams is such that the preceding models overestimate the effects of a national team player learning and improving more after he joins an elite club.

In the preceding models, it is clear that players who are categorized as being on elite and non-elite clubs were exposed to that kind of "team treatment" during the year. However, consider a case where a national team player joins his first-ever elite club in July of year t. This may be a relatively common occurrence since the top European leagues commence in late August and so the summer of year t would coincide with the start of training for a new club

[20] The key instrument in these equations is a Bosman dummy which equals one in 1996. See note 19 supra for the derivation of this form. One might prefer a longer Bosman time period dummy as an instrument. When we use either a two- or three-year post-Bosman dummy in the column 2 or 5 models that considers longer windows of years, the estimated coefficients on the PCTELITE variable in these 2SLS models is virtually unchanged in magnitude or level of significance.

season. If this player helped his national team to an improved record in games from January to June of year t, and the national team's record improved no further in the games from July through December, then the improved national team performance reflects selectivity by the elite club and not training and learning from the player's exposure to the elite club. This section estimates national team performance models that consider this issue.

One way to address this possibility would be to look at the effects of the prior year club affiliations of national team players. However, the www.national-footbal-teams.com data source on national teams does not report the club affiliation in year t-1 for players on national teams in year t but who were not on the national team in year t-1. An alternative method which we can implement with our data is to examine the effects of changes in a national team's PCTELITE variable from year t-1 to year t on changes in ELO points accumulated from July 1 of year t through June 30 of year t+1. Note that, if players do improve from any new exposure to elite clubs, models using this new definition of the dependent ELO team performance variable should underestimate any such learning effects. For example, if a player has been on an elite club for all or most of the year and helps his national team improve early in year t because of this new exposure to an elite club, then the ELO point value of the team is being raised by learning effects due to the elite clubs prior to July 1. This in turn reduces the change in ELO one would measure from this higher level of ELO on July 1 of year t through to June 30 of the following year.

Table 10 reports estimates of PCTELITE effects from various models where the change in ELO points from July 1 of year t through June 30 of year t+1 is the dependent variable and the change in PCTELITE is measured from year t-1 to t. We directly compare results from models reported in the previous tables to the analogous models with the revised measure of change in ELO points. In each pair of columns in Table 10, the results in the (a) columns refer to the previously reported results in which the change in ELO points is measured from January 1 to December 31 of year t, and the (b) columns show results for models in which the change in ELO points is measured from July 1 of year t through June 30 of year t+1. Table 10 compares the estimated effects of the change in PCTELITE variable across the two different ways of measuring the timing of the change in ELO Points in six different samples and specifications: (1) full sample OLS model of change in ELO on change in PCTELITE (Table 3, column 1); (2) OLS model of change in ELO on change in PCTELITE across pairs of years where roster remains at least 80% the same (Table 3, column 6); (3) OLS model of change in ELO on percent of a roster that will be first time elite club players next year and the percent of a roster that are first time elite club players this year (Table 6, column 1); (4) OLS model of change in ELO on change in PCTELITE from the year before to the year after the Bosman ruling (Table 9, column 1); (5) two-stage regression discontinuity model in a narrow window of years just before and just after the Bosman ruling (Table 9, column 4); and (6) two-stage regression discontinuity model for the full sample of years (Table 9, column 2).

Table 10: Effect of Change in Elite Players on Change in National Team Elo Ratings

(Annual Change in ELO Rating from January 1 to December 31 and from August 1 to July 31; Model specifications match those from the associated tables identified in Row 3.)

	1a[a]	1b[b]	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
Δ in % Elite	66.17*** (0.000)	14.65 (0.417)	107.86*** (0.003)	86.60** (0.040)	--	--	103.0* (0.100)	175.19** (0.023)	276.5*** (0.004)	191.48** (0.019)	1214.0*** (0.000)	461.12*** (0.000)
% Roster who are 1st time Elite next year	--	--	--	--	67.14*** (0.001)	19.88 (0.400)	--	--	--	--	--	--
% Roster who are 1st time Elite this year	--	--	--	--	112.59*** (0.005)	78.97** (0.046)	--	--	--	--	--	--
N	1766	1754	278	276	1707	1695	86	86	230	229	1766	1754
Jan. to Dec Model Preview table	Table 3 column 1	—	Table 3 column 6	—	Table 6 Column 1	---	Table 9 column 1	—	Table 9 column 4	—	Table 9 column 2	—
Sample Description	Full Sample 1990 - 2010		% Same Roster > .8 1990 - 2010		Full Sample 1990-2010		1995 – 1996 Pre-Post Bosman Change Only		1993 – 1994 to 1995 – 1996 IV Estimate		1990 to 2010 IV Estimate	

Table Note: (a) Models in a columns Measure Change in ELO Rating from Dec 31 to Dec 31
(b) Models in b columns Measure Change in ELO Rating from July 31 to July 31

The results in the Table 10 models show that in most cases, as expected, the magnitude and level of significance declines between the original estimates in the (a) columns and the analogous new estimates in the (b) columns. Further, except for model 1b, the full sample OLS model of change in ELO, for the estimated PCTELITE effects in the (b) columns where the models likely underestimate any learning effects from players joining elite club teams, the effects of the PCTELITE variable remain positive and significant. If we focus on the coefficients in columns 4b and 5b that focus on estimates of PCTELITE right around the time of the Bosman ruling from fuzzy regression discontinuity specifications, the coefficients there imply that a one standard deviation difference in the change in PCTELITE variable leads to a change in ELO of about 12 points, which in turn would move a national team up by roughly 4 ranking spots at mean ELO rankings. Moreover, if one examines the estimates in column 3b that compare the effect of first time elite players in the year before and the year after they become elite club team members, one observes an insignificant effect in the year before players join their elite clubs and a large positive significant effect in the year after they join their elite club for the first time.

VIII. Player-Level Data and Samples

The preceding models provide strong evidence of significant treatment-of-the-treated and local average treatment effects of the PCTELITE variable on team level performance. The measure of performance we use in these national team-level models is an especially convincing one designed for this particular setting and arguably the most consequential measure in soccer – did the change to an elite club by one or more players on the national team help the national team win? In this section, we examine the question about the effect of elite team membership on performance with unique data on player-level performance outcomes.

Player-Level Sample and Data

To analyze the effects of changes in a player's club affiliation on his individual performance requires player-level performance data. Data on individual player performance in soccer is not nearly as detailed as it is in many other sports. However, for this study, we obtained detailed proprietary data collected by OPTA Sports, Inc, a private sports data collection firm, that allows us to develop player-level performance metrics. OPTA data measure all "events" in a soccer game (passes, shots, corners, headers, throw-ins, etc) and characteristics of those events (players making the pass or taking the shot, location of the pass or shot on the field, opposing team, time stamps for each event, and so on) for a large number of soccer games of professional soccer clubs around the world, especially European clubs, as well as for many of the leading international team competitions, especially the World Cup and the Euro Cup.

OPTA data on European league club games and international team games are available for the 2007 season (fall 2007 to spring 2008) through the 2010 season (fall 2010 to spring 2011) season, or a total of four years. The number of games and teams covered in 2007 is relatively small and more games are covered by OPTA in each successive year. Because data for the analysis in this section on player performance in national team games comes from OPTA data as well, we measure the national team's performance over a time span for games from August 1 through July 31 of year t that corresponds closely to the time span for the seasons of elite European clubs. We therefore relate player-level performance in national team games from August 1 of year t through July 31 of year t+1 and relate it to club affiliation listed for year t. Using these data on game-specific soccer "events," we construct the following player-level performance metrics based on a player's passes, shots, and goals for four different twelve month periods:

1. *Pass Completion Percent*: Percent of passes completed by a player during a year

2. *Player's Passing Skill Fixed Effect*: Person specific coefficient in a regression of a Pass-Completed Dummy on (a) a dummy for identity of opposing team, (b) a set of dummy variables for the starting and ending locations of passes to control for the difficulty of the pass,[21] and (c) person fixed effects. These person fixed effects are calculated for each year a player's data are available. We use a player's fixed effect as his measure of passing skill and ability. It measures how much more or less likely any player is to complete a pass, controlling for the difficulty of completing the passes he makes and the identity of the defending team.

3. *Player Forward Passing Fixed Effect*: Same metric as metric 2, but measured for a sample of passes travelling forward.

4. *Shooting Percent*: Simple goals per shot ratio. The sample for this analysis is restricted to players with a minimum of 10 shots in a year.

5. *Player Shooting Fixed Effect*: Player specific coefficient from a regression of the probability that a shot will be a goal on (a) the identity of the opposing team and (b) the distance, distance squared, angle, and angle squared of the shot (as measured by the polar

[21] We consider six possible starting and ending location of a given pass such as right side of back third to right side of middle third, or pass starting and ending within the right-hand side of the final third. Dummies for a richer set of grids would of course be desirable but this number is limited by the need to have enough passes to and from any pair of grids so that this model can be estimated.

coordinate distance and angle of the shot using the center of the goal as the origin for calculating polar coordinates).

6. *Expected scoring*: This metric measures the average probability across all players that a shot taken from the specific polar coordinate will be a goal. This metric assesses whether a player is taking shots from spots on the field that, for an average player, are more or less likely to be successful. This can be considered a decision-making metric.

Players are included in this sample of player-level data if they have 2 or more consecutive years of international game data (over the 4 year period of OPTA data) with at least 50 "events" in each year. There are 816 individuals of which 360 have 2 years, 371 have 3 years, and 85 have 4 years of data, for a total of 2173 player-year observations. Players also must have a minimum of 10 shot events in each year to be included in the models that calculate their shooting fixed effect metric. Since shots are relatively rare events, this is a small sample of 36 players and 86 player-year observations (with 3 players having 4 years, 8 players having 3 years, and 25 players having 2 years of data).

IX. Models of the Effects of Elite Club Membership on Player-Level Performance

Table 11 presents the coefficients on the variable "elite," (a dummy for whether the player was on an elite club in the season that matches the twelve month period for which national team performance is measured) in the regressions using the various player performance metrics as dependent variables. Controls included in these models are: the previous year's performance measure, season, and player fixed effects. Thus, in these models with controls for player-fixed effects, the coefficient on the ELITE variable measures within-player changes in passing or shooting performance between national team periods when the player is and is not on an elite club. We exclude any observations from the sample for any player-year that does not have data in a contiguous year to ensure that the player and national team members have otherwise changed as little as possible.

According to the estimates in column 1, a given player is 2 percentage points more likely to complete a pass in the season after he switches onto an elite club—an increase of about 1/5th of a standard deviation. In column 2, one observes that after a season when the player switches onto an elite team, he is 2 percentage points (about ¼ of a standard deviation) more likely to complete a pass, controlling for the difficulty of the pass he is attempting and the quality of his opponent; while in column 3, this estimated effect increases to about 3 percentage points (almost 1/3 of a standard deviation) when the sample only includes (presumably more difficult) passes that travel forward. Joining an elite team improves passing performance in a significant and meaningful way.

Table 11: Effect of Playing on Elite Team on Player Performance in National Team Play
(All models include fixed effects by person and a lagged value of the dependent variable, and standard errors are clustered by person.)

	(1)	(2)	(3)	(4)	(5)	(6)
	% completed passes	Individual fixed effect in pass regression[a]	Individual fixed effect in forward pass regression[a]	Goals per shot	Individual fixed effect in shot regression[b]	Probability of scoring given origin of shot[c]
Elite	**0.021*** **(0.061)**	**0.021**** **(0.023)**	**0.029**** **(0.038)**	**0.137**** **(0.013)**	**0.181**** **(0.012)**	**-0.027***** **(0.001)**
Season	0.005* (0.0586)	-0.005** (0.0239)	-0.005 (0.111)	-0.014 (0.553)	-0.012 (0.676)	-0.005 (0.454)
Constant	-9.182* (0.088)	10.15** (0.024)	9.851 (0.112)	27.27 (0.553)	24.01 (0.677)	10.80 (0.446)
N	1,357	1,357	1,357	49	49	49
R-squared	0.137	0.181	0.175	0.183	0.226	0.423
# of players	816	816	816	36	36	36

a. The dependent variable is the player specific coefficient in a fixed effects regression run on all passes in a given season where the dependent variable is a dummy for whether the pass is completed and controls include opponent, path of pass, and starting position,

b. The dependent variable is the player specific coefficient in a fixed effects regression run on all shots in a given season where the dependent variable is a dummy for whether a goal is scored and controls include distance, $distance^2$, $distance^3$, angle, $angle^2$, $angle^3$ and dummies for right angle and for the right side of the field.

c. The probability is derived from a mean of predicted values for a player's shot coordinates in the regression described above but estimated without player fixed effects.

Robust p values in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Columns 4-6 present results on the effect of being on an elite team on shooting performance. The results must be interpreted with care, even though the estimated effects of elite club membership on these shooting performance metrics are all significant, since there are only

36 individuals who have data on over 10 shots for at least two consecutive years. Still, according to the column 4 model, joining an elite team increases goals per shots by 14 percentage points (1.4 standard deviations) and increases the probability of scoring given the coordinates of the shots' origin by 18 percentage points (2 standard deviations). Interestingly according to estimates from the column 6 model, being on an elite team increases the difficulty of shots as the coordinates of the shots taken point to a lower probability of scoring. For this very small number of players, getting on an elite team makes them more likely to score but also more likely to attempt more difficult shots.

Results from a series of models that use as dependent variables six different advanced player-level performance metrics provide strong and consistent evidence corroborating the team-level performance models. When a player becomes a member of an elite club, his national team's performance improves right away. A key reason for this effect according to the estimates shown in Table 11 is that, when a player joins an elite soccer club, that player's own soccer performance improves more than does the performance of other players.

We go further with this individual level data to explore the effects on those national team players who never play on elite teams. We examine whether their pass performance in national team play is affected by the percent of national team members that they play with who are on elite teams.[22] Table 12 presents the results. According to row 1, all measures of passing performance, controlling for past performance, are higher in teams with more elite players. However in row 2 where the independent variable of interest is change in percent roster elite, we do not find significant effects. Therefore while we do find evidence that the change in performance of less talented players is higher in teams with more elite players, we do not find evidence that the change in performance of never elite players is higher in teams where the change in percent elite is higher. Still the row 1 results are consistent with the idea that there is skill transfer from the elite players to the non-elite players when they return to play on their national teams. Not only are these players better because of their elite experience but their teammates also become better.

[22] We cannot examine shooting performance since very few of the never elite players have enough shots to be included in the shooting analysis.

Table 12: Effect of Teammates on Elite teams on Never Elite Players' Performance in National Team Play
(Controls included lagged values of the dependent variable.)

	Dependent Variable					
	Pass %	Pass %	Pass FE	Pass FE	F-Pass FE	F-Pass FE
	(1)	(2)	(3)	(4)	(5)	(6)
% Roster Elite	0.054***		0.073***		0.102***	
	(0.009)		(0.000)		(0.000)	
Δ in % Roster Elite		0.007		0.006		0.046
		(0.881)		(0.874)		(0.434)
Observations	643	643	643	643	643	643
R-squared	0.469	0.464	0.341	0.325	0.274	0.255

Robust p values in parentheses
*** p<0.01, ** p<0.05, * p<0.1

X. Interviews with Professional Players: Evidence on Mechanisms Behind the Peer Effects

The preceding sections offer clear and consistent evidence of positive, significant, and meaningful productivity effects that individual players enjoy when they work with a team of elite peers. The evidence comes from multiple sources – detailed personally-constructed team-level and player-level data sets. In this section, we add a third category of evidence to the analyses – interviews with a select group of national team players. The purpose of the interviews is not only to offer additional evidence that such peer effects exist, but also to provide a richer understanding of mechanisms behind those effects. Short interviews of approximately 30 minutes each were conducted with six former U.S. national team players. All six players were members of the national team during the 1990-2010 sample period considered in the team-level models. All players interviewed played on several clubs during their careers. All six spent considerable time on European clubs and five were members of clubs classified as elite in the preceding econometric models. The number of caps these players have earned ranges from approximately 30 to over 100.[23]

The interviews were open ended conversations but all six interviews focused on two questions designed to explore the experiences of these players without suggesting the nature of the results from the team- and player-level models. The opening question was:

"We are interested in the development and improvement of national team players.

[23] The authors thank Sunil Gulati for arranging the interviews.

Your national team career went from [year 1 to year 2] and you earned [##] caps in that time. Can you think of one or two periods during your national team career when you felt like you improved the most – when you had a breakthrough in your performance? We are not necessarily interested here in the times when you felt you played your best soccer for the national team, but periods with especially big improvements in your performance. If you can identify any examples of those kinds of periods, what do you think was responsible for the rapid development?"

The second question focused attention on their club experience:

"How did your different club experiences affect your development as a soccer player and a member of the national team?"

In the interviews around this second question, follow up questions focused on getting specific examples that explained any effects that membership on clubs had on these players.

Three of the five players who played on European clubs classified as "elite" in the previous empirical models identified their moves to these clubs as periods of especially noticeable growth in their games in their answers to the first question.

"I'd have to identify the time right after I went to [Big 5 League]. … Every weekend, I was playing against the best competition. And playing against the best competition every week will force you to improve." (Player 2)

"One time was in the period I played for [Champions League club X]. I was surrounded by the best players around. That really accelerated my development. … I knew if I wanted to play here, I had to work to get my game to a whole new level. … It was a new level of training, a new level of competition, a new level of professional pressure. This just kind of shot me forward…. Really, it was the surroundings I was in, the pressure I was put in day in and day out from coaches, teammates, and fans. … Then, in the national team play in that period, you could be more of a natural influence on the team and the games. " (Player 4)

"I have to say that the period from [year 1 to year 2] was the highlight. That's when I was with [Big 5 League club X]. Playing with and against the best players

in the world was something special. You learn so much from the others around you. That's when you seem to improve the most." (Player 6)

The other three players also confirmed both the importance of learning from teammates on top European teams and the idea that the elite club teams accelerate development.

"When I went to Europe [to a second division club in a Big 5 League country], I played at a high level for 38 league games and definitely found that the game slowed down for me. It was the first time the game became slower than my thinking. There is much more pressure in the better leagues. More pressure to perform in practice so you get a spot in the game. More pressure to perform in the game to keep the spot. More pressure from the press. You constantly need to perform." (Player 5) [24]

"Players have to improve to compete at higher levels [on an elite club]. A player's thought process has to get faster and you have to go through that. Every little thing matters. You have to work in much tighter spaces and that means more speed. You see players get to the right spaces more quickly. It's all related and all of it means a capability to do all those things faster." (Player 4)

"Sure, playing on a top division European team will raise your game. The second you raise the level of play, you have the best opportunity to improve. Day in and day out, you have the speed of the game being faster, the speed of opposition. The speed of decision making has to improve." (Player 1)

Especially interesting in these conversations with U.S. men's national team players were the various mechanisms they identified as explanations for the improvements they experienced from playing on top European teams. Most observations reflect peer effects, but the substance of what the peers were able to demonstrate to their new teammates was diverse. Certainly, several

[24] Player 5's switch from a non-elite to an elite club, as classified in this study, came when his second division team was promoted to the top division. His comment on breakthrough improvement here concentrates on joining the second division team and focuses on his exposure to the club during that period. While the comments suggest ways to improve the definition of "elite" clubs, the comments certainly echo the importance of peer effects from elite teammates.

comments reflected soccer skill and knowledge as suggested in several of the preceding comments.

> *"Your first touch is a good example. On other teams, you can get away with not having as good a first touch. You can have touches that go a little too far away. This has to improve since you'll lose the ball under higher level competition pressure. You can reach out to a coach who will fire hundreds of balls at you, but then that improvement in drills continues to get tested and reinforced in scrimmages and games. Then when you improve there, you have more time for yourself on the ball which can lead to better decisions, better passes. You just learn more, like looking around and behind you before you receive a ball. Maybe you knew that before, but now you have to make these habits to play at this level. ... If you are the best player on your team or league, you are going to plateau. So if that's happening for a player at a lower level league, he won't be challenged anymore." (Player 3)*

While all players interviewed echoed these sentiments about technical skills, another theme stressed by the players was learning from their peers about how to deal with the pressure of high-level soccer on and off the field.

> *"You also see that it's how you live and your attitude toward how you spend your time away from the training so you're ready to train. ... Right next to me, I had [list of four internationally- acclaimed internationals from other countries]. I saw right in front of me how it was done. They were mentors right in front of you to watch." (Player 4)*

> *"You learn in this kind of environment how to deal with the media. You learn a new language. You learn how to play with other players with a different language than your own. You're an American guy over there and learn how to manage that. ... How you live as a soccer player – all those things interact to make you a better player." (Player 2)*

> *"There are some other things too. The media and culture requires acclimation. The British media isn't shy about pointing out your shortcomings and that's not the*

case in other cultures with other leagues. You get used to the exposure and criticism." (Player 3)

Perhaps expected, two players singled out their coaches on elite teams as especially important "peers."

"My coach [at Champions League team X] had a big influence on my learning. He taught me a lot about how you think in my position. You get much more detailed information from that kind of coach. It was an incredible learning experience. [He] had been national team coach of the [top ranked] national team and other countries' teams and clubs. He taught me an enormous amount."(Player 4)

"My coaches [on this English elite team] were expecting different things. I had to change daily habits, eating habits, and so on." (Player 6)

The players also warned us against thinking that the elite peer teammates were teachers in the same way a coach might be.

"Your peers are now guys who are looking for maybe that last 2 year contract. … [My team] was in the Champions League then. There were players from the Scottish, Dutch, and Turkish national teams all at my position. We all were trying hard to get better than players who were already at a very high level. Of course, we all wanted the playing time. It was very competitive." (Player 4)

"It wasn't easy. You'd really be getting stuck in during practice some times."[25] (Player 6)

The player interviews also offered some cautions about the main relationship investigated in this study between elite club membership and accelerated productivity improvements. First, the interviews cautioned us against concluding that this relationship would happen for every

[25] "Getting stuck in" refers to hard physical play characterized especially by tough tackles aimed at communicating a competitive attitude toward the opposition.

player. Related to this first point, some players also highlighted other factors that may be needed for players to benefit from what elite teams have to offer.[26]

> *"The flip side is that I can see how this [effect of playing with elite peers] wouldn't happen. Your confidence can take a shot. You were succeeding and now everything is happening so much faster. Going from [a youth national team] to [a Champions League team in a Big 5 League] was a massive step." (Player 3)*

> *"There are exceptions. [A player on the team I now coach outside of the Big 5 leagues] trains at such a high level of intensity every single moment that he makes that higher level environment for himself."[27] (Player 1)*

> *"But I need to caution you, there is an effect of a guy going to Europe that alters people's perception of that player. You get a cache. ... Some players who play in World Cups, they get there because of that cache. That's not necessarily the same thing as better play or player." (Player 2)*

Two players concluded that the peer effects from playing on elite club teams were very real but that there are some additional hard-to-measure characteristics that elite club players tend to have that help them benefit from their elite team peers.

> *"So many players from the US are going over there now. It's really great. But I am worrying some because it's not just what and who you are on the field. It's being able to go into and adapt to a whole new culture, and just being a professional. So many are now going over there at a very young age. Players could certainly struggle. They need to know that it's not always going to go well. How you deal with that and work through that is important." (Player 4)*

---

[26] The results from the econometric models certainly do not contradict the points to follow. We do not have the data necessary to estimate coefficients on our Elite or PCTELITE variable that vary by player or team. Even though we find very consistent evidence throughout for productivity enhancing effects of playing on elite club teams, the experience of some individual players or teams could of course not follow this overall pattern.

[27] In a follow up question, we asked if this player would or would not still accelerate his improvement by going to an elite club in a Big 5 league. The response was that "Yes, he'd improve. If you keep playing with better players, you just have a chance to play at that higher level and improve."

"I'd be shocked if these [peer] effects didn't happen for everyone [from MLS clubs or from the U.S. U20 national team], as long as they were mentally strong enough to deal with all of the pressure." (Player 5)

XI. Conclusion

Large sample studies of individual workers have long interpreted positive effects of tenure on wages or productivity as evidence of on the job learning. However, the mechanisms behind increases in worker productivity as experience increases are rarely studied. This study investigates whether the talent of an individual's co-workers helps explain differences in the rate of human capital accumulation on the job. When an individual works as part of an elite team, does his productivity increase faster than when his teammates are less talented?

With data on professional soccer teams and players that are particularly well suited for developing convincing non-experimental evidence about these kinds of peer effects, the empirical results consistently show that performance improves more after an individual has been a member of an elite team than when he has been a member of lower level teams. The conclusion is borne out by a rich set of complementary data on: national team performance, player-level performance, performance of foreign players who joined elite teams after an exogenous shift in the number of foreign players participating on top club teams, performance of players on national teams in the year just before and the year just after they join an elite club team, and experiences of several national team players obtained through personal interviews. An increase of one or two players on a national team joining an elite club can lead to substantial improvement in national team play that can change the world ranking of the team by several spots.

While the unique data on soccer players and teams are important for addressing the difficulties in estimating productivity effects due to high talent peers, the conclusion that team assignment and team talent levels will change the productivity of peers will not apply to every production process. However, where such effects exist, productivity effects due to peers can be long lasting and build from one period to the next, thus leading to high and low performing groups. Within a given firm, team assignment and composition can therefore be important managerial decisions. Hiring high talent workers has spillover effects. While this study adds to the new literature on peer effects in the production process by estimating peer productivity effects that are longer lasting and that transfer across organizational boundaries, this study also lends added support to the broader conclusion that real production functions are very complex and very challenging to manage.

References

Abdulkadiroglu, Atila, Josh Angrist, and Parag Pathak (2014) “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools,” *Econometrica,* vol. 82, no. 1, pp. 137-196.

Bandiera, Oriana, Iwan Barankay, and Imran Rsaul (2005) “Social Preferences and the Response to Incentives: Evidence from Personnel Data” *Quarterly Journal of Economics*, vol. 120, pp. 917-962.

Dahl, Gordon B., Katrin V. Loken, and Magne Mogstad (2014) “Peer Effect in Program Participation,” *American Economic Review*, vol. 104, no. 7, pp. 2049-2074.

Dobbie, Will, and Roland G. Fryer, Jr. (2014) “The Impact of Attending a School with High-Achieving Peers: Evidence from New York City Exam Schools,” *American Economics Journal: Applied Economics,* vol. 6, no. 3, pp.58-75.

Duflo, Ester, Pascaline Dupas, and Michael Kremer (2011) “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *American Economic Review*, vol. 101, pp. 1739-1774.

Falk, Armin, and Andrea Ichino (2006) “Clean Evidence on Peer Effects,” *Journal of Labor Economics*, vol. 24, no. 1 (January), pp. 39-57.

Frick, Bernd (2009) “Globalization and Factor Mobility: The Impact of the ‘Bosman Ruling” on Player Migration in Professional Soccer,” *Journal of Sports Economics*, vol. 10, no. 1, (February), pp. 88-106.

Ichniowski, Casey and Kathryn Shaw (2012) “Insider Econometrics," in *Handbook of Organizational Economics*, ed. by R. Gibbons, and J. Roberts. Princeton University Press.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote (2012) "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees," *American Economic Review*, vol. 102, no. 5, pp. 2048-82.

Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez. (2013) "Taxation and International Migration of Superstars: Evidence from the European Football Market" *American Economic Review*, vol. 103, no. 5, pp.1892-1924.

Lazear, Edward P., Kathryn L. Shaw, Christopher T. Stanton (2012) “The Value of Bosses” NBER Working Paper No. 18317.

Mas, Alexandre, and Enrico Moretti (2009) “Peers at Work” American Economic Review, vol. 99, pp. 112-145.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses" *American Economic Review*, vol. 103, no. 4, pp. 1289-1324.

Sacerdote, Bruce, 2001 “Peer Effects with Random Assignment: Results from Dartmouth Roommates,” *The Quarterly Journal of Economics*, vol. 116, no 2, pp. 681-704.