

Extreme wages, performance and superstars in a market for footballers

by **Rachel Scarfe, Carl Singleton and Paul Telemo**

[Discussion Paper](#) No. 2020-04

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

Extreme wages, performance and superstars in a market for footballers*

Rachel Scarfe[†]

University of Edinburgh

Carl Singleton[‡]

University of Reading

Paul Telemo[§]

University of Edinburgh

November 2020

Published at: *Industrial Relations: A Journal of Economy and Society*,
60(1): 84-118. (2021). <https://doi.org/10.1111/irel.12270>

Abstract

We study the determinants of superstar wage effects, asking whether productivity or popularity-based explanations are more appropriate. We use longitudinal wage and performance data for workers (players) and firms (teams) from a particular market for sports talent: Major League Soccer (MLS) in the United States. We find evidence that the top earners, whose annual salaries are mostly not accounted for by their past MLS performances, when compared alongside other footballers, are paid more because they attract significantly higher stadium attendances and thus revenues. There is no evidence that higher residual salary spending by the teams affects their relative performance in football terms, or that the amounts the teams spend on actual talent affect attendances. Taken together, these results suggest that a popularity-based explanation of superstar wage effects is appropriate among the top earners in this labour market.

Keywords: Superstar effects, Top incomes, Labour productivity, Major League Soccer

JEL codes: D31, J24, J31, L83, Z22

*We are grateful for helpful comments and advice from two anonymous reviewers, Dennis Coates, Maia Güell and J. James Reade, as well as from participants at the 2019 SGPE Conference (Crieff, UK). We also thank Leon Singleton and J. James Reade for their research assistance. The data used in this study are publicly available from Major League Soccer (MLS) LLC and the MLS Player's Association. The use of these data does not imply the endorsement of the data owners in relation to the interpretation or analysis of the data.

[†]School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh, EH8 9JT, UK; rachel.scarfe@ed.ac.uk

[‡]Department of Economics, University of Reading, Edith Morley Building, Reading, RG6 6UB, UK; c.a.singleton@reading.ac.uk; with thanks to the Economic and Social Research Council (UK) for funding support under Grant No. ES/J500136/1

[§]School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh, EH8 9JT, UK; paul.telemo@ed.ac.uk

1 Introduction

Some people have labour incomes so high that they are barely comprehensible to the average worker. These top earners, and the amounts they receive, attract widespread media, political and academic interest.¹ Over the past forty years and throughout the major developed economies, the differences between the wages of most workers and the few at the top have risen, especially in English speaking countries. For example, in the United States, the share of total income earned by the top 1% of earners more than doubled between the 1970s and the 2010s (Alvaredo et al., 2013). In addition, an increasing share of the wealthiest individuals in society have earned rather than inherited their wealth (Kaplan and Rauh, 2013). As a result, inequalities in labour income are an increasingly important driver of overall inequality. However, there is no clear consensus about why some individuals earn such enormous amounts or the causes of these ‘superstar wage effects’.

Top wage earners include business and finance executives, lawyers, medical professionals, and media and sports stars. We focus on the latter group, using longitudinal data on the salaries and productivity of both workers (players) and firms (teams) in US Major League Soccer (MLS), between 2007 and 2018. This is a specific group of workers, but it is nonetheless relevant to overall patterns in wage and income inequality. Media and sports stars are among the professions with the fastest growth in top incomes: the media and sports stars in the top 0.1% of US earners experienced a 5% compound annual growth in their incomes between 1979 and 2005 (Bakija et al., 2012). They are also well represented among the wealthiest individuals around the world. For example, Franzini et al. (2016) report that two-thirds of the 120 top wage earners in Italy in 2003 were association footballers. In contrast to most other labour markets, there are detailed, accurate, linked, firm and worker productivity data in sports labour markets (Szymanski, 2007). We combine these with salary data, which are available for all players in the League due to MLS regulations. Because of this data availability and the clarity of the relevant rules and institutions, markets for sports talent provide a natural and convenient setting to answer a simple question: why do some individuals earn such astronomically high wages?

We use a two-step empirical strategy to approach this question. In the first step, we use past individual player performance data, such as the number of shots on target per 90 minutes on the football pitch, to decompose annual MLS salaries into two parts: (i) the amount explained or predicted by on-the-pitch performance or productivity; (ii) an unexplained or residual amount. We hypothesise that the latter part may capture the ‘superstardom’ of some MLS players. In the second step, we regress measures of teams’ annual football performance or revenue generation on the aggregate predicted and residual wages of their players, as estimated from the first step. Across various model specifications and robustness checks, we find consistent evidence that the amount teams spend on the residual wage component increases their revenue positively and

¹Note the extensive coverage of the various Forbes rich lists, e.g., for sportstars: <https://www.forbes.com/athletes/list/>. Income taxes for the highest earners are often a subject of debate in national elections, e.g., in the December 2019 UK general election (see Adam et al., 2019).

significantly. A 1% increase in the residual salary spending of teams increases home stadium attendances by as much as 0.14%. This team-level spending on the residual salaries of players is not significantly associated with a team's relative performance in MLS, including whether they make the end-of-season playoffs. Conversely, the spending on the predicted part of player wages does significantly increase a team's points per game and chances of reaching the MLS playoffs, whereas the residual salary spending does not.

We relate these results to the two main theories of superstar wage effects, which are discussed in more detail in Section 2. One theory, proposed by Rosen (1981), suggests that in markets where there is a convex relationship between productivity and revenue, small differences in talent can result in large differences in wages. In contrast, Adler (1985) posited that superstar wages can result from differences in popularity rather than talent. Our results are more supportive of Adler's theory, as we find that teams do not perform better on the football pitch when they spend more on superstars, but they do attract more people to their stadiums. This suggests that the top wages in MLS are a reflection of player popularity, rather than of current talent or performances on the football pitch.

Our main contributions to the existing literature are twofold. First, we add to the debate on the causes of superstar wage effects. Our data allow us to both link a worker's wages with their individual performance (or productivity) and to accurately measure firm productivity over time, and thereby distinguish between productivity- and popularity-based theories of superstar wage effects. Another benefit of our data is that we observe approximately the universe of MLS players over the period we study, allowing further investigation by weighting each player by the time they spent on the football pitch in a season. Our results are robust to this weighting. This suggests that teams benefit from the popularity of their superstar players even when they are not playing that much, adding further support to Adler's theory of superstars. Second, we contribute new evidence about the determinants of pay in MLS and football more generally, as well as the impacts on teams of signing superstars. For example, by demonstrating the positive relationship between a team's spending on residual wages and its revenue, we are able to suggest a source for the large amount of unexplained variation in football player wages.

The remainder of the paper proceeds as follows: Section 2 discusses the related literature on superstar wage effects; Section 3 describes the estimation strategy and interpretation; Section 4 summarises the institutional setting of MLS and the data sources; Section 5 presents our results; and Section 6 concludes.

2 Related literature

2.1 Theories of superstar wage effects

Rosen (1981) defined the presence of superstars in a labour market as the "concentration of output among a few individuals, marked skewness in the associated distributions of income and very large

rewards at the top." The two principal and competing theories to explain how some individuals become superstars and earn vastly greater income than most others were those proposed by [Rosen \(1981\)](#) and [Adler \(1985\)](#). The former's theory states that superstars exist in markets where there is a convex relationship between productivity and revenue. This convexity arises because of the imperfect substitution between individuals of different productivity or talent, and because technology allows the joint consumption of output. In the music industry, for example, consumers prefer better musicians and are unwilling to substitute with others whom they consider to be inferior. At the same time, musicians can reach large audiences at a low cost, through technologies such as TV and the internet. Thus, small differences in talent can lead to large differences in revenue, and hence very large salaries for the most talented. In contrast, [Adler \(1985\)](#) argued that superstar wages can occur even without differences in individual talent. In his framework, there can be a number of equally talented individuals. However, consumers then derive utility from consuming the same product as others do (e.g., from listening to the same musician), and from knowing more about a particular individual. This knowledge is gained in discussion with other fans, so that as a musician becomes more popular, consumers prefer them even more. At the same time, it is costly for consumers to search for and identify the "very best" musician. Therefore, it is optimal for consumers to listen to the most popular musician, even when there are others, slightly more talented but less popular. In this way and unlike in [Rosen's](#) theory, superstars can emerge by chance, simply because they are initially slightly better known than others or through opportune timing, rather than because of small differences in talent. Beyond these two main theories, [Terviö \(2009\)](#) suggests an alternative explanation of superstar wages, which may be relevant in markets where workers and firms are not able to commit to long-term wage contracts, and where talent is only revealed through actual performance on-the-job. In these markets, [Terviö](#) shows that firms will excessively bid for known talent rather than trying out new talent. This is because, although talent may not be scarce, the supply of workers whose talent has been revealed is scarce.

If superstar wages in football can be explained by [Rosen's](#) theory, then we would expect to see small differences in individual productivity leading to differences in firm (team) revenue and large differences in player wages. If [Adler's](#) theory provides a better description of superstar wages in football, then we would not expect such a clear relationship between individual productivity and revenue. Instead, we would expect that differences in an individual's popularity, measured as something unrelated to observed productivity, can explain differences in revenues and wages. In a particular football talent market featuring superstars, namely MLS, we can observe different measures of individual player and team productivity and revenues, and can therefore test whether the [Adler](#) or [Rosen](#) theories, or some combination of the two, are more relevant. We are generally silent on whether the [Terviö \(2009\)](#) theory is an appropriate description of superstars in football, other than to note that the players and teams can commit to long-term contracts, even when players are very young (e.g., [Simmons, 1997](#); [Frick, 2007](#)) and that talent is normally revealed to teams through extensive scouting and coverage of player performances.

2.2 Evidence of superstar wage effects

There are a number of practical difficulties in identifying why some individuals attract superstar wages. First, in most settings it is difficult to objectively measure an individual's productivity and relate this to wages. Some authors have attempted to use other metrics as proxies for productivity. For example, [Hamlen Jr \(1991\)](#) used a singer's harmonic voice quality, and found that the elasticity of record sales with respect to voice quality is positive but small. He concluded that this suggested the superstar effects described by [Rosen](#) do not exist in the music industry. Similarly, [C  lerier and Vall  e \(2019\)](#) used the results of university entrance exams in France as a proxy for Chief Executive Officer productivity. They found that the returns to this measure were higher in finance than in other industries, and they interpreted this as evidence of superstar effects in the finance industry. However, these measures of talent are highly specific, and may miss other dimensions of productivity. As well as productivity, it can also be difficult to measure very high earnings accurately. Few people are required to publicly report their salary (sports labour markets, particularly in the US, are an exception), and data are often top-coded.²

A second problem is found in disentangling measures of productivity from measures of popularity. For example, [Krueger \(2005\)](#) used the amount of space devoted to musicians in an encyclopedia of music as a metric for "star quality", and found that this measure was associated with an increase in an artist's revenue. This measure was related to both talent and productivity, and it was not possible to identify which drove the higher revenue. Likewise, [Hoffman and Opitz \(2019\)](#) analysed superstar effects in the film industry. They showed that measures of popularity, such as internet search hits, positively affected the earnings of highly paid film stars. They interpreted this as evidence of "publicity superstars", as described by [Adler](#). However, if those film stars had previously been nominated for an Oscar, then their earnings were not related to popularity. They suggested that this group are "talent superstars", as described by [Rosen](#).

Finally, it can be difficult to identify whether superstar effects are indeed a result of exogenous technologies that allow large numbers of people to consume the same output (a key component of both the [Adler](#) and [Rosen](#)'s theories). Instead, the adoption of new technologies may be endogenous, and driven by the presence or emergence of superstars within a market. [Koenig \(2019\)](#) tested for the mechanism creating superstar effects in the US labour market for entertainers. With a plausible source of exogenous variation in the adoption of a new technology, coming from when and where local TV stations launched, he showed that wages at the very top of the income distribution increased markedly due to the ensuing increase in audience sizes, and that the incomes and employment of less talented entertainers decreased.

²Some studies, e.g., [Hamlen Jr \(1991\)](#); [Krueger \(2005\)](#), have used revenue (in these cases, from music sales) as a dependent variable, but do not observe exactly how much of this revenue is paid to the artists. Others, e.g., [C  lerier and Vall  e \(2019\)](#), have used self-reported data from questionnaires.

Evidence of superstar wage effects in sport

Professional sport provides a setting to evaluate theories of superstar wages as both detailed productivity and wage data are available for both firms and employees. The literature exploiting this fact has focused on two main questions. First, a number of studies have asked what determines the wages of sports superstars. For example, [Lucifora and Simmons \(2003\)](#) found that the wages of Italian footballer players are highly convex in two particular measures of performance, namely the numbers of assists and goals scored per 90 minutes, suggesting that small differences in talent are indeed multiplied into large differences in wages, consistent with the [Rosen](#) theory. [Franck and Nüesch \(2012\)](#) considered the effects of both talent and popularity on player values in the German football transfer market. Their measure of player popularity was the residual from a regression of media citations on measures of a footballer's talent. They interpreted their main findings, that both talent and popularity have positive effects on player values, as indicating that some footballers are superstars in both the [Rosen](#) and [Adler](#) senses. Also studying the wages in European football, [Carrieri et al. \(2018\)](#) showed that talent, popularity and bargaining power are all associated with higher wages. This association is stronger at the top of the wage distribution, and only popularity is significant at the 95th percentile of the wage distribution. The authors concluded, therefore, that a theory of superstars based only on talent differences is not adequate in this setting.

There is a second strand of research on the effect of superstar sportspeople on team (firm) revenue. In particular, there are several papers showing that the presence of superstars increases attendance and TV revenues in cricket ([Paton and Cooke, 2005](#)), baseball ([Ormiston, 2014](#)), basketball ([Hausman and Leonard, 1997](#); [Jane, 2016](#)) and football ([Brandes et al., 2008](#); [Lawson et al., 2008](#)). Particularly relevant to this paper are studies on the effects of so-called 'designated' players on the league results and attendances of MLS teams ([Coates et al., 2016](#); [Jewell, 2017](#)). These designated players can be hired outside the salary cap that normally applies to teams. They tend to be the highest earners in MLS by some margin, with teams often signing well-known but ageing players, who were coming to the end of highly successful careers in the major Europe leagues and international competition (this is colloquially known as the 'Beckham' rule, since David Beckham was among the first beneficiaries). [Jewell \(2017\)](#) found that only a few of the designated players signed since the rule was introduced in 2007 had positive effects on stadium attendances (namely Beckham, Blanco and Marquez), that these effects diminished over time, and that they were larger when the superstars played away from their teams' home stadiums. In other words, these players generated a novelty factor in the League, which spilled over beyond the teams who signed the players. [Coates et al. \(2016\)](#) found that, whilst teams with a higher wage bill after the introduction of the designated player rule did perform better in the league, this was offset by a decrease in performance for teams with higher salary inequality among players. Although these studies suggested that superstar players do increase team performance and attendance (and hence revenues), they did not try to distinguish whether increased attendance was due to the players' superior talent or their greater popularity. Attempting to disentangle these mechanisms, [Bryson et al. \(2014\)](#) considered the effect of migrant players in the top tier of Italian football on both league points and attendance. They found that teams with more migrant players performed better

in the league and had higher attendances. Using wage data, they also found evidence that these migrant players could be thought of as superstars in both the productivity-based (Rosen) and popularity-based (Adler) senses.

3 Empirical strategy

To measure how the allocation of playing talent in MLS affects a team’s output, we use a two-step regression methodology. Our empirical strategy is similar to that used by Bryson et al. (2014), except here we introduce a manner of weighting the influence of each player in their respective team, potentially allowing us to dig deeper into whether players are superstars in the productivity- or popularity-based senses. In brief, in the first step we regress the contracted wages of football players on several measures of their performance during the past season and some other observable characteristics. We treat the predicted part of this regression as the amount of the wage that can be explained by factors relevant to a player’s productivity on the pitch. The residual component of these regressions, which is by construction orthogonal to the productivity component, is the estimated amount of the wage which isn’t accounted for by what players individually achieve on the pitch. In other words, it may reflect remuneration for some other contribution to the team’s output, such as from the popularity or ‘superstardom’ of the player. We then aggregate these player-level measures to the team level for each season. In the second step, we regress the output of teams on the derived aggregates from the first step, which ought to reflect what their players are being paid for talent or for other unobserved contributions to their team.³ In doing so, we look for evidence as to why some players are paid especially high wages in MLS, i.e., superstar wage effects.

We consider two principal measures of an MLS team’s output, Y_{jt} : (i) a team’s performance on the pitch (represented by its results in the League); (ii) its ability to generate revenue (represented by attendance at its home games). Consider the following general representation of the output of football team j in season t :

$$Y_{jt} = F_j \left(P_{jt}, Z_{jt}, \tilde{Y}_{jt}, t \right), \quad (1)$$

where F_j is the team-specific production technology, P_{jt} is the quantity of playing talent, and Z_{jt} is the amount of superstardom of the team’s players. \tilde{Y}_{jt} represents other contemporaneous measures of output, which may affect Y_{jt} . For example, it is plausible that a team’s success in the League, i.e., winning matches, would encourage higher stadium attendance and revenue.

³We make several other changes to the method used by Bryson et al. (2014). For example, in the first step, we estimate wage regressions separately by primary playing position, as the returns to different productivity measures may significantly differ in this way, i.e., between defensive or offensive players.

To measure the quantity of playing talent and superstardom employed by a team, we assume that the log wages of player i in season t are given by:

$$w_{it} = G(p_{it}, z_{it}, t), \quad p_{it} \perp z_{it}, \quad (2)$$

where p_{it} is the level of talent on the pitch and z_{it} is some uncorrelated other factor affecting wages. As with the team's output, player wages are affected by League-wide trends. We estimate Equation (2) using least squares:

$$\ln(w_{it}) = \mathbf{x}'_{it-1} \beta_k + d_{kt} + z_{it}, \quad (3)$$

where \mathbf{x}_{it-1} is a vector of lagged performance variables and personal characteristics, reflecting the importance to the team as a footballer, e.g., experience and whether they are the team captain, and β_k is the associated vector of coefficients. We allow the effects of each variable to vary by the player's primary position, where $k = K(i) = \{\text{Goalkeeper, Defender, Midfielder, Forward}\}$. To allow us to decompose player wages into a part paid in respect of playing talent, and a part paid in respect of other factors, it is important that the performance variables, \mathbf{x}_{it-1} , relate only to performance on the pitch. For this reason, we do not include characteristics such as a player's nationality, which has nonetheless been shown to relate to pay (e.g., [Bryson et al., 2014](#); [Thrane, 2019](#)). We lag the performance variables since player contracts and wages are predetermined at the beginning of a season. This has two econometric benefits. First, it will provide more confidence in the second-step regression, since it is then not the case that the player performance measures (e.g., scoring goals) used in the first step are mechanically affecting the teams' current season outputs in the second step (e.g., winning football matches). Second, it helps to address the potential concern of reverse causality in the two-step model, which could occur if a team's current season output directly impacts the amount of talent they are able to hire. d_{kt} are season-position fixed effects and the remaining heterogeneity in player wages is in the residual term, z_{it} . This could include heterogeneity due to player popularity.

To generate estimates of P_{jt} and Z_{jt} , we aggregate within a team and season the estimates from Equation (3) as follows:

$$\widehat{P}_{jt} = \sum_{i \in (j,t)} \widehat{p}_{it} = \sum_{i \in (j,t)} \omega_{it} \exp(\mathbf{x}'_{ikt-1} \widehat{\beta}_k), \quad (4)$$

$$\widehat{Z}_{jt} = \sum_{i \in (j,t)} \omega_{it} \exp(\widehat{z}_{it}), \quad (5)$$

where $i \in (j, t)$ denotes the players who are in team j in year t . ω_{it} is the weight of each player in the overall team-level aggregates. Thus, \widehat{P}_{jt} is the part of the team-level aggregate wage bill that is explained by player performances, and \widehat{Z}_{jt} is the part that is not explained by performances, and which may reflect 'superstardom'.

In our baseline estimation, we give each player on a team’s roster in a given season the same weight, i.e., $\omega_{it} = 1$. We also consider whether players’ contributions to their team’s output depend on the amount of time they actually spend on the football pitch. To do so, we compare the baseline with results where player i ’s weight in year t is given by:

$$\omega_{it} = \frac{m_{it}I_{j(i)t}}{\sum_{i \in (j,t)} m_{it}}, \quad (6)$$

i.e., the share of the overall number of minutes played by all players on the team’s roster that season, multiplied by the total number of players in the team and season, I_{jt} . Players who did not feature at all during the season would be given zero weight. This gives a low weight to young and inexperienced players, who play few minutes, and to players who are injured or suspended, even if they are highly paid. Any difference with the baseline results may tell us something about the mechanism by which superstardom affects a team’s output, and may suggest whether the [Rosen](#) or [Adler](#) theories are more accurate in this setting. We also consider other values for ω_{it} , such as only giving weight to positive or very high values of \hat{z}_{it} .

To relate these aggregate components of a team’s salary spend to its output, as per Equation (1), we estimate the following regression model:

$$\ln(Y_{jt}) = \alpha_j + \lambda \ln(\hat{P}_{jt}) + \gamma \ln(\hat{Z}_{jt}) + \mathbf{X}'_{jt} \delta + d_t + v_{jt}, \quad (7)$$

where α_j are team-specific fixed effects, such that we identify the model using variation over seasons in the output within teams. We are primarily interested in the coefficients λ and γ . These provide estimates of the elasticity of this season’s output to the estimated overall quantities of playing talent and superstardom within teams, respectively. \mathbf{X}_{jt} contains other relevant time-varying factors related to team output, with δ being the associated vector of coefficients. We also consider time fixed effects in the model, d_t , and v_{jt} is the residual.⁴

The team output measure and model interpretation

In practice, we begin by using team performance on the pitch as the dependent variable in Equation (7). If we find that the total measure of residual wages across all players on a team, \hat{Z}_{jt} , does not affect a team’s relative on-pitch performance, i.e., $\hat{\gamma}$ is small and insignificant, this suggests that we have indeed captured in \hat{Z}_{jt} a component of wages which is unrelated to how successful the team is in purely football terms. This would provide confidence in our decomposition of wages into a part explained by player productivity and a part that may reflect ‘superstardom’. In effect, some players would be paid wages that are higher than we would expect

⁴ We also estimate this equation including the lagged output variable on the right-hand-side, using two-step General Method of Moments (GMM), based on the approach suggested by [Arellano and Bond \(1991\)](#). This is to address concerns about reverse causality; teams with previous high attendance may be able to pay higher wages. We find that the lagged output variable is insignificant in all our model specifications, and so we omit it and estimate using OLS.

based on their individual productivity, and these additional wages would not seem to be paid in respect of any unobserved characteristics that improve their teams' relative on-pitch performances.

We repeat the second step of the model, using the average home stadium attendance of teams as the dependent variable. If we find that \hat{Z}_{jt} positively affects a team's revenue generation, i.e., $\hat{\gamma}$ is large and significant, then it suggests that some players are paid especially high wages in MLS because of their popularity, and that teams (and the League) can financially benefit from this. This would be more consistent with [Adler's](#) theory of superstars rather than that of [Rosen](#). In particular, if we also find that the aggregate level of playing talent on the pitch, \hat{P}_{jt} , only increases revenue through the team's relative success in the League, and not in its own right, i.e., $\hat{\lambda}$ is small and insignificant, then this would provide further evidence supporting [Adler's](#) theory. Conversely, if we find that a team's aggregate playing talent positively affects revenue generation, this suggests that consumers (football fans) prefer to watch the most talented players, which would be more consistent with [Rosen's](#) theory.

4 Data and institutional setting

Major League Soccer is a talent market with several unique features. In this section, we discuss the relevant regulations in MLS, set out our data sources, and provide some descriptive statistics on player wages.

4.1 MLS institutions and regulations

MLS is different to the most popular association football leagues in Europe or elsewhere in a number of ways. MLS teams compete in two parallel closed leagues (i.e., no relegation or promotion): the Eastern and Western Conferences. In each season (calendar year), teams play each other team in their conference twice, and each team in the other conference once, known as the regular season. They earn points for winning or drawing a game, following which the top six teams in each conference advance to the MLS playoffs. These 12 teams play a direct elimination (knockout) series to determine the championship winner, known as the MLS Cup. Separately, the team with the highest regular season points across the two conferences is awarded the Supporters' Shield. Thus, both a team's points and whether or not it advanced to the playoffs are relevant measures of team performance

Unlike most other football leagues, MLS is a single-entity that owns a stake in all the teams, which are run as franchises. The teams receive some direct revenues, including local broadcast rights, all stadium revenue and 70% of match day revenues, such as ticket sales ([Peeters, 2015](#)). They also receive a share of the overall league's profits, including from national and international broadcast rights and league-wide sponsorship.

Salary regulations

Of particular importance to this study are the salary rules in MLS. Players sign a contract with the League, rather than with an individual team. This limits individual players' bargaining power, and has kept teams' salary costs low relative to their revenue, compared to other football leagues around the world (Twomey and Monks, 2011). There are lengthy rules governing salaries and which players a team can sign. These are subject to Collective Bargaining Agreements (CBAs) between the League and the MLS Players' Association (MLSPA). There were three CBAs during the period that we study, covering the periods 2004 to 2010, 2010 to 2015 and 2015 to 2020. Negotiations between the MLSPA and MLS have been fractious, as players have argued for higher salaries and more choice in which team they can play for.

The CBAs specify a minimum annual salary for every player and a total salary cap for every team.⁵ These salary floors are binding for a small percentage of players. For example, in 2018, 26 players aged at least 24 were paid the minimum salary, out of a total of 537 in the League. Players earning the minimum salary are generally a homogeneous group: younger, new entrants to MLS, and often play no or few minutes during a season.

The salary cap is the total budget that teams have for all players on their roster, i.e., who can play for them during the season. Teams generally have the freedom to choose how to allocate wages within the cap, subject to minimum and maximum roster sizes, as well as some rules about the composition of player types on the roster, including international vs home-grown players. As MLS has become more successful, the MLSPA has been able to negotiate substantial increases in both the minimum salary, from \$30,000 in 2007 to \$70,250 in 2019, and the salary cap, from \$2.1m in 2007 to \$4.2m in 2019. There have been other changes during the period that we study, including an increase in job security. Both the 2010 and 2015 CBAs increased the percentage of players with 'guaranteed contracts'. These are contracts that cannot be terminated by MLS if the player performs badly or is injured during a season. Thus, every player will spend at least one full season in MLS. At the same time, the number of 'option years', where teams can extend a player's contract for an extra year, was reduced (Ferrari and Rueda, 2015).

There are two main ways that teams can spend money on salaries in excess of the cap. Teams can sign a number of players using the 'designated-player' rule. This was introduced in 2007 when David Beckham, who was at that time one of the highest paid footballers in the world, signed for LA Galaxy from Real Madrid of the Spanish La Liga. Until 2009, the rule allowed each team to sign one player whose salary did not count toward the cap. In 2010, this was expanded to two designated players per team, with an option of paying a fee for a third.⁶ The rule was designed to allow teams to sign high profile players from outside leagues, at salaries that would not be feasible under the salary cap (Coates et al., 2016). Teams also receive 'allocation money' from the League each season, which can be used to pay salaries above the cap but below the very large (essentially unregulated) salaries of designated players. This money is aimed at improving the quality of

⁵There is also a separate minimum salary for young players, below the age of 24, on a team's reserve roster.

⁶This fee is shared among the teams that do not have three designated players.

players in MLS, by increasing the overall amount spent on salaries, and at maintaining competitive balance in MLS, by allowing the younger or lower performing teams to purchase higher quality players ([Major League Soccer \(2017a,b\)](#)). MLS has a large amount of discretion in how much allocation money is given to each team. As a result, there is substantial dispersion in teams' overall salary costs. The percentage of players from outside the US and Canada has increased over time, from 32% in 2008 to 47% in 2018, as higher salaries have allowed MLS to attract more players from overseas. To account for the average effects of these changing regulations on player wages, we include season controls in our first- and second-step regressions.

Roster regulations

There are lengthy rules determining which players a team can sign and the make-up of its roster (see [Major League Soccer, 2020](#), for details). For example, there is a maximum number of international players that a team can include on its roster. There is also a 'draft' system where teams can pick young players new to MLS in reverse order of the team's finishing position in the League in the previous year. In general, players have little control over which team they play for. The limited form of 'free agency' introduced in the 2015 CBA is an exception, whereby players aged over 28, who have played at least 8 seasons in MLS, can negotiate with any team when their contract expires.⁷ By allowing players to negotiate wages directly with a team, and teams to compete to sign a player, free agency may increase player salaries. However, the number of players eligible for free agency during the period we analyse was small, ranging from 28 players in 2015 to 39 in 2020. There are a number of other ways that teams can acquire players and circumnavigate the rules governing team rosters and the salary cap. For example, they can trade players or their international roster allowance with other teams.

The League's structure and regulations are designed to ensure that it remains competitive. This appears to have been successful; the MLS Cup was won by 12 different teams in the 13 seasons between 2007 and 2018. This illustrates a further advantage of studying MLS: in this setting, without a small number of entrenched teams at the top of the League, managerial decisions, including on salary and the make-up of the team, can potentially have almost immediate effects on points and success, since the gaps between the teams are small and not persistent. A further advantage is that most players are still subject to the salary cap, although the teams with more resources can certainly afford to spend more on designated players. This implies that the richest teams in the League cannot always employ all the players they would wish to and cannot pay them as much as they would like to, easing concerns later about there being possible reverse causality between teams paying superstar salaries and performance.

4.2 Data sources

Our data primarily come from two main sources, the official MLS website and the MLSPA, and cover all players and teams in MLS from 2007 to 2018. Online Appendix Table [A1](#) contains

⁷The 2020 CBA has expanded this to any player aged over 24 with at least 5 seasons in MLS.

further information on all the variables used in our analysis. As our analysis relates player wages to lagged performance variables, we use data on player performance from 2007 to 2017 and on wages and team performance from 2008 to 2018.

Wage data

Player wage data and team affiliations come from the MLSPA. The data refer to the players on a team's roster midway through the MLS season in August of each year, after the secondary transfer window when teams can sign players from abroad. The player performance data for each season were extracted from the MLS online database in September 2018. We merge these two sources of data together using player names and seasons, creating a dataset containing 6,194 player-season observations, representing 2,186 different players and approximately the universe of those contracted to MLS during 2007-18. After dropping the small number of observations where players had missing records for their age or season performance, or because records could not be matched between the two data sources, and focusing on 2008-18, our sample contains 5,458 player-season observations, representing 1,939 different players.⁸ The wage measure we study is the guaranteed annualised compensation or salary, henceforth referred to as wages.

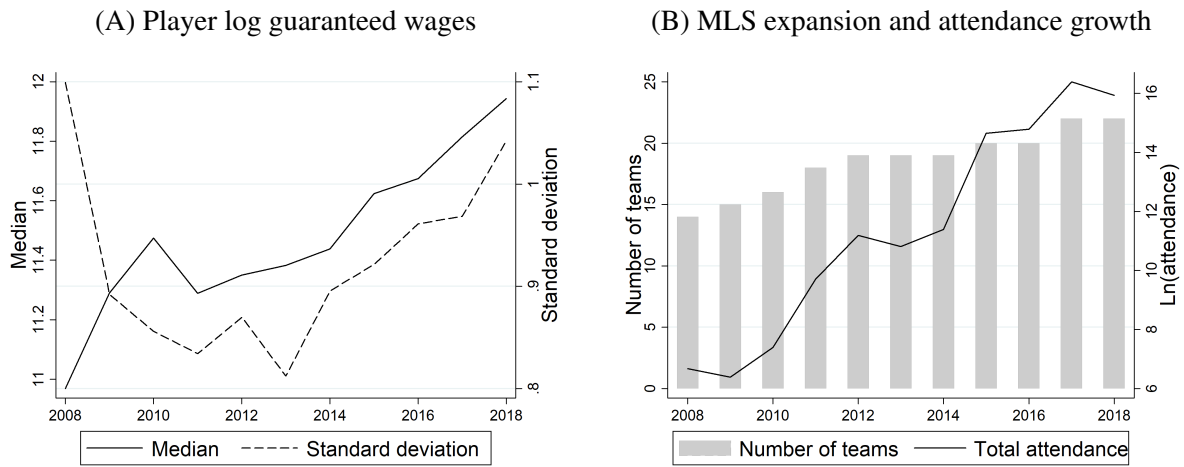
Our wage measure includes annualised payments for signing with a new team (loyalty bonuses) or related to marketing, but does not include performance-related pay. Although we do not have data on individual performance bonuses, we note the following: first, teams receive a bonus from the League for good performance, which is shared among all players in the team; second, bonuses are paid at the end of the season, while our data cover the wages agreed before the season; and third, according to MLS rules, any "readily achievable" individual bonuses are included in our wage measure for the purposes of calculating the contribution of each player's wage to the overall team-level salary cap. The total team wages that we observe (excluding any designated players) are generally higher than the salary cap (see Figure 3 below). Although the rules for calculating exact contributions to the salary cap are complicated and require information which we are unable to access, this suggests that individual performance bonuses are not large compared to the wages that we observe.

The 2007-18 period was one of expansion in MLS and growth in wages, as shown in Figure 1. Wage dispersion also increased during this time, partly due to the designated player rule (Figure 1A). As we use wage data from 2008 onward, Figure 2 summarises 'real' log wages in 2008-18 by primary playing position on the pitch, for the analysis sample of players, with wages adjusted to 2018 MLS-US\$, to address average wage inflation over the period. Although the rules regarding wage determination are complex, there is nonetheless substantial variation. The variance is higher for forwards and midfielders, who are more likely to earn a very high wage

⁸There were some discrepancies when we merged the datasets, due to naming conventions and spelling, which we manually corrected as far as possible. There were a tiny number of obscure cases where we could not reconcile the two datasets, or where players appeared in one source but not the other. See Online Appendix Table A2 for the number of players in the sample by year and position.

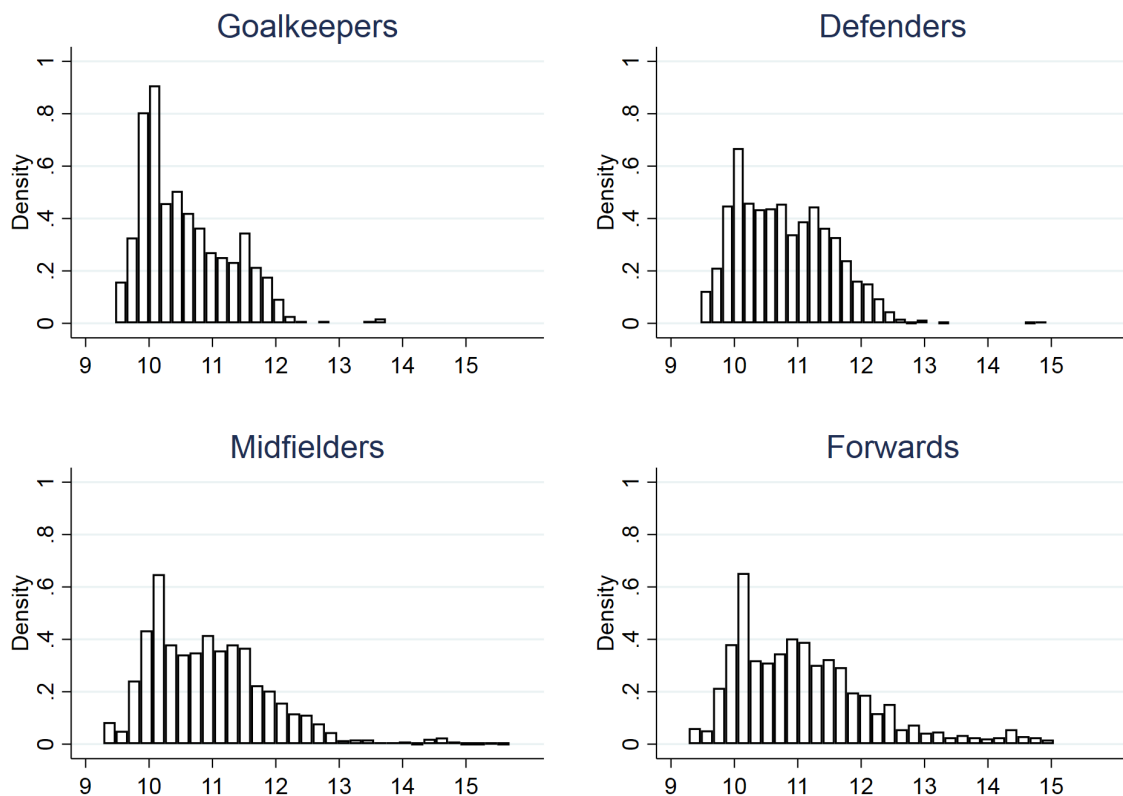
than defenders or goalkeepers. Designated players make up 7% of the player-season observations, with some players switching to or from this status during their MLS careers.

FIGURE 1: MLS growth and expansion: player wages, teams and attendance, 2008-18



Notes.- Player wages are nominal US\$.

FIGURE 2: Distribution of MLS log guaranteed wages by player position, 2008-2018

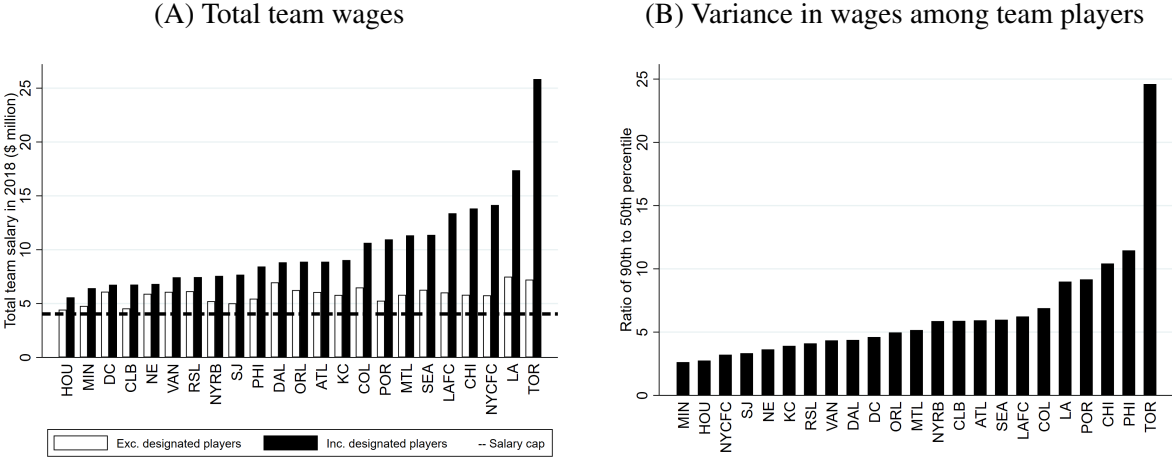


Notes.- 2018 MLS-US\$: to address general wage growth and inflation within the MLS when pooling over years, the wages of all players in each year were adjusted according to the average wage level in 2018, i.e., they are roughly adjusted to something akin to 2018 MLS prices.

Although there is a salary cap, in practice it appears that teams are able to spend very different amounts and make very different decisions regarding wages. Panel A of Figure 3 shows the

total wage bill for each team in 2018. All teams spent more than the salary cap, with the extra coming from designated players and allocation money. The largest sum was spent by Toronto FC, nearly five times as much as Houston Dynamo. The wage dispersion between players in the same team also varies substantially, as shown in Panel B, which plots for each team in 2018 the ratio of player wages at the 90th and 50th percentiles. To illustrate this heterogeneity over teams, the largest wage bill and highest variance in wages at Toronto FC in 2018 were due to three designated players (Sebastian Giovinco, Michael Bradley and Jozy Altidore), who together earned over US\$18million, and were respectively the 1st, 2nd and 7th highest paid players in the League in 2018.

FIGURE 3: Team wages in 2018



Notes.- author calculations using MLS Player’s Association data. Wages are nominal US\$.

Player-season-level performance data

Table 1 summarises the player-season-level performance variables used in our analysis. For forwards, midfielders and defenders, we observe the number of goals, assists and shots on goal in the previous MLS regular season, i.e., excluding the playoffs. We also observe the numbers of red and yellow cards received, and the numbers of fouls committed by and against each player. To normalise and compare between players who spend different amounts of time on the pitch, we convert all these variables into the average per 90 minutes played in the previous regular season. For goalkeepers, we observe saves as a percentage of the shots on goal by the opposing teams.

One concern is that our performance variables are more relevant for forwards and midfielders than for defenders and goalkeepers. However, we can observe the minutes played in each regular season. Assuming that team managers choose players from their roster based on performance, this should provide a good proxy for individual productivity; better players will normally play more minutes, notwithstanding injuries.⁹ There is substantial variation in minutes played during a season. On average, players are on the pitch for 44% of the total time that their team plays in

⁹This is potentially a strong assumption, as managers may wish to give more popular players more time on-the-pitch, regardless of their productivity, to keep fans happy.

a regular season, and the standard deviation is 29%. Another indicator of a player’s productivity or value to the team, which we expect to affect wages, is whether they are the team’s captain for the current season. We use data on historical team captains from Wikipedia, which we spot check against news articles and individual team websites.

We construct three dummy variables indicating new entrants to MLS, re-entrants, who were not signed by a team in MLS in the last season but who had previously played in the League, and transfers, who were traded to a new team in MLS between two consecutive seasons. Player turnover in MLS is high; 28% of our player-year observations are for players new to the League, and 14% are for players transferred to a new team. These variables may provide further useful information about player productivity. In particular, after a player’s contract expires, the team for which he currently plays has the first right of refusal to re-sign him ([Major League Soccer Players Union, 2015](#)). If they choose not to do so, then this may indicate that he has performed below their expectations. Conversely, a highly productive player might be transferred to another team for a high price. We also construct a dummy variable indicating a player who tends to play in other positions besides his primary role, as this versatility could make him a more productive member of the team. This applies to 7% of the player-year observations.

TABLE 1: Summary of player-season-level variables

	Mean	St. Dev.	Min.	Median	Max.
<i>All players: current season variables (2008-18)</i>					
Wages (US\$1,000s)	235.43	620.44	12.90	100.00	8,650.00
Actual age	25.91	4.45	15	25	42
Captain	0.03	-	-	-	-
Designated player	0.07	-	-	-	-
MLS new entrant	0.31	-	-	-	-
MLS re-entrant	0.03	-	-	-	-
MLS transfer	0.13	-	-	-	-
<i>All players: lagged season variables (2007-2017)</i>					
Mins played (%)	44	32	0	43	100
Multi-position	0.07	-	-	-	-
<i>Defenders, Midfielders, Forwards: lagged season variables (2007-2017), per 90 minutes played</i>					
Assists	0.11	0.14	0.00	0.07	0.65
Goals	0.12	0.18	0.00	0.04	0.79
Shots	0.43	0.49	0.00	0.23	2.12
Shots on goal	1.17	1.11	0.00	0.81	4.86
Red cards	0.01	0.03	0.00	0.00	0.18
Fouls committed	1.04	0.78	0.00	0.92	4.00
Fouls suffered	1.18	0.82	0.00	1.11	4.80
<i>Goalkeepers: lagged season variables (2007-2017), per 90 mins played</i>					
Saves	45%	33%	0%	64%	100%

Notes.- For dummy variables (Captain; Designated player; MLS new entrant; MLS re-entrant; MLS transfer; Multi-position) the mean is the share of observations equal to one. See Online Appendix Table A1 for variable definitions and sources. See Online Appendix Table A2 for the number of players in the sample by year and position.

Team-season-level data

MLS expanded during the period we analyse, from 13 teams in 2007 to 22 teams in 2018 (see Online Appendix Table A3). Table 2 summarises team-level variables. We use the average points per game played in the regular season of the League, taken from the MLS website, as our main team performance variable.¹⁰ We also construct a dummy variable indicating whether a team reached the playoffs in each season, as an alternative measure of success, and a dummy variable indicating a team new to MLS, i.e, an expansion franchise.

As explained above, the team operators are all joint investors in the League as a whole, and receive a share of the overall profits. Most team operators are individuals or private companies, and thus data on this particular component of their earnings are not available. However, operators do receive some direct revenues, in particular a majority share of any revenue derived from attendance at their team's games, such as ticket sales, parking, food and drink, and from local broadcast rights and local sponsorship revenue. The remainder of these revenues, as well as national broadcast rights, league-wide sponsorship etc., are distributed among all investors in the League (US Court of Appeals, 2002). We can, therefore, use attendance as another measure of team output. We check that attendance is not constrained by stadium capacity, which would lower the correlation between attendance and revenue, by comparing each team's average home attendance to the capacity of its home stadium (see Online Appendix Figure B1). In general, stadium capacities are substantially greater than average attendances, suggesting that team operators are rarely capacity constrained.

We use two further measures of team output in our robustness checks: team revenue and team value, estimated and published by the Forbes media company (Smith, 2013 to 2018). Forbes uses a range of information to construct these estimates, including attendance data, sponsorship deals, investments in stadiums, and broadcasting deals. Although they are only available for five seasons (2013 and 2015-18), they provide a cross-check for our results using average attendance as a measure of output. There is substantial variation in both these estimates. In these years, estimated team revenue ranged from \$13m to \$63m and estimated team value ranged from \$64m to \$330m. The Pearson correlation coefficient between annual average home attendance and estimated revenue is 0.69, and between attendance and value it is 0.57, providing reassurance that average attendance is a justifiable measure of team output.

5 Results

5.1 First-step regressions

We estimate the wage regression given by Equation (3) over the period 2008-18, separately for goalkeepers, defenders, midfielders and forwards. The results are shown in Table 3, with one column for each of these four playing positions. The fit of the estimated wage equations is low,

¹⁰Three points are awarded for a win, one point for a draw (tie) and zero for a defeat.

TABLE 2: Summary of team-level variables

	Mean	St. Dev.	Min.	Median	Max.
<i>Current season variables (2008-18)</i>					
Points per game	1.37	0.30	0.47	1.41	2.09
Attendance (10,000s)	1.94	0.71	0.71	1.82	5.30
Expansion team	0.05	-	-	-	-
Revenue (US\$millions)*	30	11	13	26	63
Value (US\$millions)*	184	68	64	175	330

Notes.- teams receive three points for a win, two points for a draw, and zero points for a loss. For dummy variables (Expansion team) the mean is the share of observations equal to one. See Online Appendix Table A1 for variable definitions and sources. See Online Appendix Table A3 for the number of teams and when they featured during the sample period.

*Nominal figures to nearest US\$million, 2013 and 2015-18 only.

with an R^2 between 0.47 and 0.62, which is in line with other studies of MLS (e.g., [Kuethe and Motamed, 2010](#)), suggesting that much of the variation in players' wages is not captured by some observable measures of their performance. Generally, there are statistically significant differences between the player positions in how the factors considered relate to wages.

Wages increase in age for outfield players (excluding goalkeepers). Unsurprisingly, the wage premium for a captain is large and significant for some positions, increasing the annual wage by 50 log points for defenders and 66 log points for forwards. For outfield players, the wage premium for a new entrant to MLS is also large and significant. However, players who return to MLS from playing elsewhere do not generally receive a significant premium. A transfer between MLS teams is associated with a wage cut of between 11 and 24 log points in the following season, which is probably a result of teams deciding not to re-sign players.¹¹

For goalkeepers (column I, Table 3), wages are concave in the percentage of minutes played in the previous season, implying that wages significantly increase until a goalkeeper plays approximately 67% of the time, and then begin to decrease after. The estimated effect on wages of the lagged percentage of shots saved is small and insignificant. For defenders (column II, Table 3), none of the lagged performance variables, including the number of minutes played, are individually significant at the 5% level in the wage regression, after clustering standard errors at the level of the player-team pair.

For midfielders and forwards (columns III and IV, Table 3), there is significant evidence that players who commit more fouls per game receive higher wages. This may relate to teams having a preference toward more aggressive players, or because fouls are associated with risk-taking or greater work effort. We also find that midfielders and forwards who suffer more fouls receive significantly lower wages on average. This may relate to the higher skill level of some players, if they are then better able to avoid tackles and challenges. The other performance variable which

¹¹Players sign a contract with a team for a minimum of one year. A player's contract may give their team the option to extend after its initial term. If the team decides not to exercise that option, then the player enters a pool of players known as the 'Re-entry Draft', from which they can be signed by another team.

TABLE 3: First-step regression estimates: performance related determinants of log guaranteed salary, 2008-18

	Goalkeepers (I)	Defenders (II)	Midfielders (III)	Forwards (IV)
Age (years - 15)	0.025	0.134***	0.076**	0.102***
Age squared	0.001	-0.002**	0.001	0.000
Captain	0.592	0.497***	0.407*	0.660**
MLS new entrant	0.013	0.341***	0.500***	0.434***
MLS re-entrant	0.125	0.150	0.317**	0.355
MLS transfer	-0.181**	-0.107**	-0.150***	-0.242***
Multi position player		0.085	0.035	-0.161
<i>Lagged performance:</i>				
Mins played (% of team)	1.429***	0.360	0.185	-0.182
Mins played (% of team) sq.	-0.709**	0.357	0.602**	1.318**
Saves (% of shots)	0.001			
Assists		-0.212	0.023	0.138
Goals		0.738	0.713	-0.003
Red cards		0.457*	3.126*	2.735
Fouls committed		0.082	0.167***	0.105**
Fouls sustained		0.027	-0.078*	-0.107***
Shots		-0.041	-0.045	0.019
Shots on goal		0.088	0.112**	0.104**
Assists squared		-0.132	1.054	0.737
Goals squared		-0.480	-1.015	0.840
Red cards squared		6.456	-21.27*	-17.49
Year fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	600	1,656	2,015	1,187
<i>R</i> ²	0.621	0.486	0.474	0.524

***, **, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, with standard errors robust to clusters at the player-team level.

Notes.- lagged performance explanatory variables are per 90 minutes played, e.g., red cards per 90 minutes played squared, besides minutes played and saves, which are percentages of total team minutes played in the season and the number of shots on target that the goalkeeper faced, respectively.

significantly affects wages is the number of shots on goal. An increase in one shot on goal per 90 minutes is associated with an 11 log point wage increase for midfielders and a 10 log point increase for forwards. Conditional on shots on target and other variables, the number of goals scored per 90 minutes does not significantly relate to wages for MLS players.

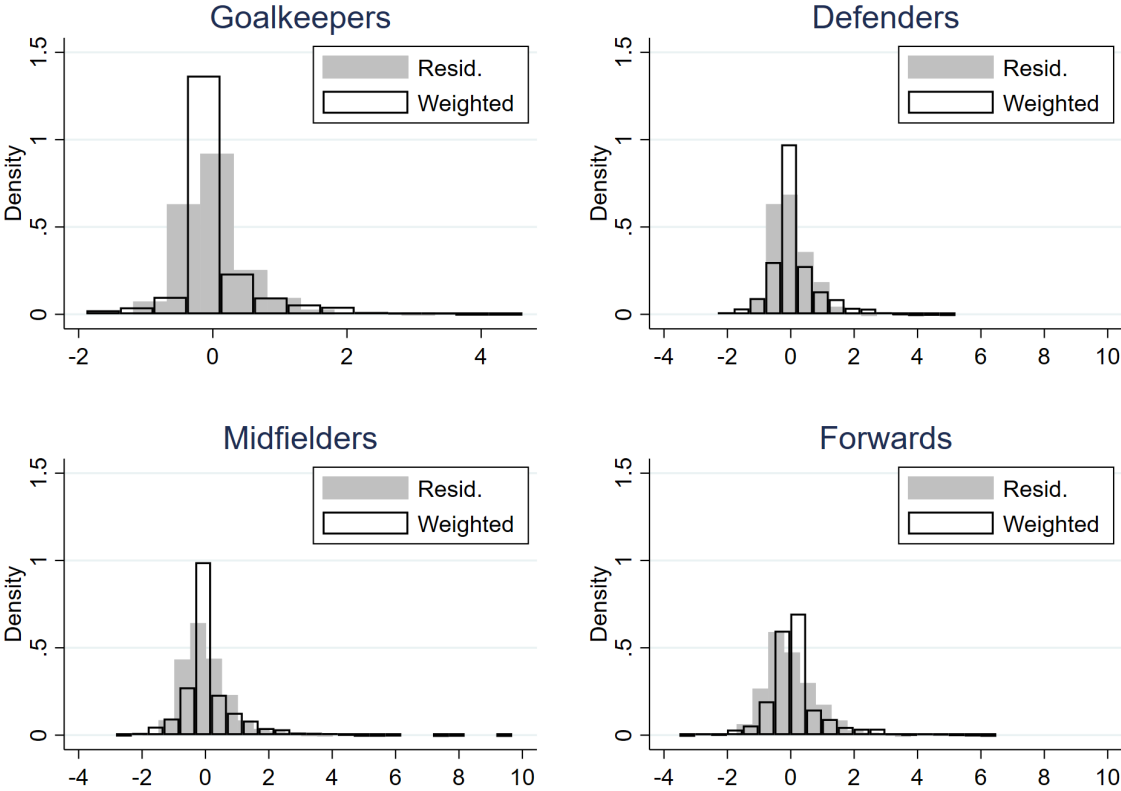
Residual wages

We proceed by collecting the predicted and residual wages from the first-step wage regressions. As Figure 4 shows, there is substantial variation in the residual wages, i.e., the part of a player's annual guaranteed salary that is not explained by measures of individual player performance, and which may correspond to superstar quality. The variation in these residuals for midfielders and forwards is greater than for defenders and goalkeepers, especially with regard to the size of the right tail. This is consistent with the variation in actual wages by position (see Figure 2). We also find that the variation is reduced when we weight each player's residual wage by the amount of

minutes they played during the season, as per Equation (6). This weighting decreases the mass of very low residual wages; players who are paid below the wage predicted by their performance also tend to play fewer minutes. This may reflect that these players have lower on-pitch productivity in ways that the wage regressions do not capture. The weighting also decreases the mass of players with very high residual wages, but who are injured or suspended during the season.

To determine whether a particular group of players could potentially drive the results of our second-step regressions, we plot the residuals for different types of player. Online Appendix Figure B2 displays the distributions of the estimated residuals, comparing new entrants with those who played in the previous season. The former are more affected by the weighting, given that new entrants generally play fewer minutes. Figure B2 also compares the estimated residuals for designated players with all other players. Designated players have considerably greater variation in their residual wages, especially after weighting by minutes played, with a generally greater mass of very high residuals in this case, i.e., designated players get time on the pitch even if their measured performances are relatively poor. Online Appendix Figure B3 shows the distribution of residuals by the region of a player’s nationality. There is a greater mass of high residuals for players from Europe and Latin America. North American players are more likely to have residuals close to zero. This is consistent with an interpretation of residual wages as being related to player popularity.

FIGURE 4: Distributions of estimated residuals from first-step player wage regressions, by position and unweighted vs weighted.



Notes.- each sub-figure plots the residuals from the regression as per Equation (3) and the columns in Table 3. The weights are described in the text.

Validity of the first-step results

To add confidence to our interpretation of residual wages as a potential measure of superstardom, we attempt to measure player popularity directly using Wikipedia page views.¹² The mean and median number of player page views per month in the sample is 2,763 and 1,027 respectively, and the most ‘popular’ player in any year was Ashley Cole, who had an average of 108,774 page views per month in 2016. From a regression of log residual wages, i.e., \widehat{z}_{it} , on log page views, a 1% increase in the latter is associated with a 0.09% increase in the former. Online Appendix Figure B4 displays the scatter plot and best linear fit corresponding to this regression, which is significant at the 1% level. To avoid omitted variable bias due to the Wikipedia page views of a player being correlated with the popularity of their team or with the league as a whole, which in turn may affect residual wages, we also report results including year and team fixed effects (Online Appendix Table C1). Adding these controls does not alter the results in any significant way. This demonstrates that a player’s popularity, as measured by Wikipedia page views, does increase his wages in a way that cannot be explained by past performances, as captured by our first-step wage regressions.

The set of on-pitch productivity regressors included in Equation (3) is limited by data availability for the whole sample period. However, since the 2013 season, more information on MLS player performances is available from the website [WhoScored.com](#). These include interceptions, attempted dribbles, pass completion rates, tackles and metres run during matches, obtained from Opta, the premier football analytics company. [WhoScored.com](#) amalgamates over 200 player performance statistics into an objective rating for every match.¹³ We obtained a season average of this rating for as many MLS players as possible. We find that the previous season’s value correlates strongly and significantly with the predicted component of the first-step wage regression, i.e., $\mathbf{x}'_{ikt-1}\widehat{\beta}_k$, though less so for defenders than other positions (Online Appendix Figure B5). We also find that a regression of $\mathbf{x}'_{ikt-1}\widehat{\beta}_k$ on a player’s season-average [WhoScored.com](#) rating, including season and team fixed effects, yields a significant slope coefficient of 0.85 and an R^2 of 0.2 (Online Appendix Table C2). This reassures us that the predicted component of the first-step wage regression is a meaningful measure of a player’s on-pitch productivity.

5.2 Second-step regressions

Tables 4–5 show the main results of the second step in our analysis, in which we regress teams’ output on their aggregate predicted and residual wage bills, i.e., Equation (7), using the results presented in Table 3 and Figure 4.

¹²Data on page views for individual players are available at a monthly frequency going back to July 2015, and were found by collecting page view data for the URL [https://en.wikipedia.org/wiki/\[player first name\]_\[player last name\]](https://en.wikipedia.org/wiki/[player first name]_[player last name]). To capture popularity at the time of contract negotiations, we use the average monthly page views in the off-season months of January and February.

¹³This is explained here: whoscored.com/Explanations.

Team performance

First in Table 4, we use team performance in the League as the dependent variable in Equation (7). In columns I-III, we use the log of average points per game in a regular season as the dependent variable, and estimate using OLS. In columns IV-VI, we use whether a team reached the MLS playoffs as the dependent variable. As this is a binary variable, we estimate the probit model equivalent of Equation (7). Columns I and IV show our results using the unweighted aggregate predicted and residual total wages for each team and season as regressors. In columns II and V, we use the predicted and residual wages from the first step for a team's designated players only, i.e., we give zero weight to non-designated players when aggregating. We do this to check whether these particular players are driving our results. Finally in columns III and VI, we use the predicted and residual wages for all players weighted by their contributions of time on the pitch, as per Equations (5) and (6). Compared to the team production suggested by Equation (1) and (7), we also omit the log of team average attendance, season fixed effects and any other explanatory variables from the estimated models. We find no significant evidence, conditional on the included team fixed performance effects, that attendance affected performance on the pitch, that there were MLS trends in points per game (e.g., fewer drawn games), or that expansion teams had better or worse than average performances in their first season.

When using all players' wages, including non-designated players, we find weak evidence that the estimated coefficients of predicted wages are positive and significant, whether using points per game or reaching the playoffs as the dependent variable. In Table 4 column I, for example, an increase of 1% in a team's aggregate predicted wages is associated with a 0.2% increase in points per game. In other words, teams that spend more on wages for talented players perform better in the League. This effect decreases slightly when using weighted wages but is more precisely estimated (column III). The coefficient on the residual wages, $\hat{\gamma}$, is insignificant. This is robust to the choice of dependent variable and to whether we weight players by time on the pitch. This suggests that the unexplained part of player wages from the first-step regressions is not associated with better team performance. It also suggests that the residuals in our first step do indeed capture a component of wages that is unrelated, or orthogonal, to player talent, and gives us confidence that we are generally not missing important unobserved elements of individual productivity, e.g., team leadership qualities, in the predicted part of wages. If residual wages reflected these qualities, then we would have expected that higher residual wages should positively affect teams' on-pitch performances. Unsurprisingly, because reaching the playoffs is based on points achieved during the regular season, we find qualitatively similar results for the probit model estimates in columns IV and VI.

In Table 4 columns II and V, we check whether the results described above are driven by the highest-paid designated players. We find that the estimated effects of residual and predicted wages for these players on team performance are insignificant. This suggests that spending more on designated players is not associated with teams doing better on the pitch in MLS. This is unsurprising, given that these players can only plausibly have small impacts on a team's

TABLE 4: Second-step regression estimates: log points per game and whether made the playoffs, 2008-18.

	Points per game			Playoffs		
	(I)	(II)	(III)	(IV)	(V)	(VI)
<i>Log wages, all players (\$m):</i>						
Predicted ($\hat{\lambda}$)	0.237*			1.938**		
	(0.122)			(0.786)		
Residual ($\hat{\gamma}$)	0.065			0.531		
	(0.081)			(0.535)		
<i>Log wages, designated players (\$m):</i>						
Predicted ($\hat{\lambda}$)		-0.003			0.301	
		(0.028)			(0.189)	
Residual ($\hat{\gamma}$)		0.025			0.027	
		(0.022)			(0.166)	
<i>Log weighted wages, all players (\$m):</i>						
Predicted ($\hat{\lambda}$)			0.221**			1.773***
			(0.092)			(0.484)
Residual ($\hat{\gamma}$)			0.060			0.490
			(0.059)			(0.432)
Team fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	204	167	204	186	142	186
<i>R</i> ²	0.262	0.288	0.279			
Log pseudolikelihood.				-110.2	-85.3	-107.1

***,**, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters [23 for Columns (I)-(III), 19 for (IV) and (VI), 18 for (V)] and displayed in parentheses.

Notes.- Columns (I)-(III) are estimated using OLS. Columns (IV)-(VI) give probit model estimates of slope coefficients.

performance; there are typically only two such players per team, who will not play every game, and who usually appear for short spells in MLS, making it harder for them to integrate with their longer-serving teammates or into tactical setups.

Home attendance

In Table 5, we use the average attendance at home games for each team and season as our second measure of output. Column I shows our base specification of Equation 7. This includes as a regressor the log points per game in the regular season, as we generally find this to be significant in explaining variation in home attendance within teams. This allows us to control for the effect of team performance on attendance; teams that win more games may attract more spectators. It also includes a dummy variable for expansion teams, as there is some significant evidence of higher attendances in a team's first year in MLS compared to later seasons.

Column I shows that aggregate predicted wages, without weighting, do not significantly predict home attendance within teams. However, aggregate residual wages do, with an elasticity estimate of 0.14. This result, when combined with the evidence we have described above, is consistent with the Adler theory of superstar wage effects. We find no evidence that aggregate residual wages predict a team’s performance on the pitch. However, these residuals positively affect home attendance, and thus a team’s revenues and profits. In other words, the results are consistent with the highest wages in MLS having little to do with the current talent of the players receiving them, but are instead a reflection of the fact that these players are popular and draw larger crowds into the stadiums. This is consistent with the previous literature that has studied superstar effects in MLS (e.g., Coates et al., 2016; Jewell, 2017), though these papers did not rule out on-pitch productivity-based explanations for very high wages at the same time.

In column II of Table 5, we estimate the same model as in column I except that we give non-designated players zero weight when aggregating predicted and residual wages from the first-step wage regressions. These results show that the residual wages of the designated players alone do not on average significantly affect home attendance. In other words, the superstar wage effects in MLS appear to be more general than just those implied by the ‘Beckham Rule’, which the previous literature has focused on (e.g., Jewell, 2017). In column III, we show estimates of the same model specification as column I, except that we use the predicted and residual wages for all players weighted by their contributions of time on the pitch, as per Equations (5) and (6). We find weaker positive effects of the team’s residual wages when applying this weighting, with an elasticity of 0.06, which is only significant at the 10% level. This is further suggestive evidence in support of Adler’s superstar wages theory over Rosen’s, in this context, as teams benefit financially from spending on superstar wages largely regardless of the amount of time that the recipients are on the pitch displaying their footballing talent.

Finally in columns IV and V of Table 5, we investigate the response of attendance to different portions of the residual wage distribution. Instead of aggregating the residual wages of all players in a team, we separately aggregate the estimated residuals from the first step of players whose residual is below and above the 90th percentile (across all seasons and player observations). The latter set of players are those who are paid much more than their previous performances would suggest and who may be ‘superstars’.¹⁴ The 90% figure is largely arbitrary, so we vary it later, but it is also consistent with there being on average one or two superstars on a team’s roster per season in MLS. We estimate the equivalent regressions as per columns I and III, except that we include the two separate measures of a team’s ‘high’ and ‘low’ residual wages and omit the overall measure. When not weighting each player’s influence in these measures by time on the pitch, we find that the coefficient estimate on the high residual wages of teams, $\hat{\gamma}_H$, is positive

¹⁴24% of these players are designated players, higher than the 6% of designated players in the full sample. This is intuitive, as it is the designated players that are not subject to the salary cap, and can therefore be paid very high wages. However, there is significant salary overlap between designated and other players. Some designated players will be paid highly in recognition of high performance and will have low residual wages. Conversely, some other players will be overpaid relative to their performance, and will have high residual wages. This measure of ‘superstardom’ captures these players.

TABLE 5: Second-step regression estimates: log home attendance (10,000s), 2008-18.

	(I)	(II)	(III)	(IV)	(V)
Log points per game ($\hat{\phi}$)	0.139** (0.065)	0.145*** (0.049)	0.131* (0.071)	0.109 (0.066)	0.114* (0.063)
Expansion team ($\hat{\delta}$)	0.106* (0.060)	0.085 (0.063)	0.126** (0.056)	0.073 (0.053)	0.153** (0.057)
<i>Log wages, all players (\$m):</i>					
Predicted ($\hat{\lambda}$)	0.166 (0.127)			0.076 (0.121)	
Residual ($\hat{\gamma}$)	0.136** (0.053)				
<i>Log wages, designated players (\$m):</i>					
Predicted ($\hat{\lambda}$)		0.026 (0.027)			
Residual ($\hat{\gamma}$)		0.020 (0.014)			
<i>Log weighted wages, all players (\$m):</i>					
Predicted ($\hat{\lambda}$)			0.133 (0.093)		0.156* (0.088)
Residual ($\hat{\gamma}$)			0.062* (0.033)		
<i>Log wages, split by the residual percentile (\$m):</i>					
Residual below 90th percentile ($\hat{\gamma}_L$)				-0.044 (0.146)	
Residual above 90th percentile ($\hat{\gamma}_H$)				0.057** (0.053)	
<i>Log weighted wages, split by the residual percentile (\$m):</i>					
Residual below 90th percentile ($\hat{\gamma}_L$)					0.011 (0.079)
Residual above 90th percentile ($\hat{\gamma}_H$)					0.028 (0.019)
Team fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
N	204	167	204	195	191
R^2	0.862	0.895	0.859	0.891	0.846

***,**, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters (21 for Column (V), 23 otherwise) and displayed in parentheses.

and significant: a 1% increase in the amount of high residual wages increases home attendance by 0.06%. An increase in a team's overall wage bill that is spent on players below the 90th percentile of the estimated residuals does not significantly affect home attendance. As with the comparison

between columns I and III, column V shows that the effect of the high residual wage measure is weaker and insignificant when we weight individual players by time on the pitch.

Online Appendix Figure B6 shows the coefficient estimates, $\hat{\gamma}_H$, obtained by re-estimating the second-step regression presented in column IV of Table 5, now varying the cutoff level of what constitutes ‘high’ residual player wages. In general, the results show that residual wage spending above the 80th percentile by teams has a positive and significant effect on their home attendances. The coefficient estimate is only slightly decreasing as the cutoff increases from the 80th to the 95th percentile. Since this should be interpreted as an elasticity, and because we are accumulating the wage of fewer players as the cutoff increases, this would imply that the absolute or relative responses of attendance to changes in residual wages are increasing with the cutoff value, i.e., players with higher residual wages have larger effects on home attendances at the margin.

In summary, we find that a team’s residual wage bill has a significant positive effect on attendance and no effect on team performance. This suggests that footballers can be superstars of the type identified by Adler (1985). The unexplained part of a player’s wages does not appear to be related to the unobserved elements of performance (or productivity) on the pitch as a footballer, but instead represents a popularity premium. This is slightly different to the findings of Bryson et al. (2014), who investigated whether the higher wages earned by migrants in the Italian football league could be explained by their superstar status. Using a similar empirical strategy, they found that migrants in Italian football earned a substantial wage premium (i.e., that the residual part of their wages was higher than for domestic players). They interpreted this wage premium as evidence of superstar wages, and found that teams with a higher residual wage bill for migrants enjoyed both better on-pitch results and higher attendance. In this paper we define a superstar as any player who earns much more than their on-pitch performance would suggest. Our findings are therefore more general: in MLS, on average, superstar wages are paid to players in respect of greater popularity, rather than higher productivity. There may exist specific groups of players in MLS, such as migrant players, who can command wage premia, or domestic players, who may suffer wage penalties, but we do not investigate this here.

5.3 Further robustness checks

To add to our confidence in these results, we perform three more sets of robustness exercises. First, we look for direct evidence that superstar effects can drive teams’ revenue (and not only home attendance). Second, we use different methods of estimating the first-step regression to reduce the impact of outliers in the data. Third, we estimate the second-step regression giving zero weight to players who play less than 20% of the time, or who are paid a minimum wage, to ensure these observations are not influencing the main results.

Other measures of revenue

So far, we have only demonstrated that superstar wage effects in MLS are consistent with an interpretation whereby players receive those amounts due to a popularity premium, increasing

home attendances and thus team revenues. However, for a small number of years we can check whether this is the only mechanism, using estimates of MLS teams' annual revenues and values. Online Appendix Table C3 summarises the results from estimating variants of Equation (7) whereby the dependent variable is the team's overall revenue or value. We include in these regressions the log of home attendance as an explanatory variable. Conditional on this, we find no significant evidence that the measures of total team revenue or value are affected by the predicted or residual wages of teams from the first step regressions. Based on only a few years of data and broad financial estimates, from a source that may not be especially reliable, this suggests that the superstar wage effects in MLS significantly relate to stadium gate receipts but not the remainder of a team's revenue or value.

Outliers in the first-step regression

The first-step wage regression of Equation (3) may be affected by outliers among players for some of the past season performance measures, because some players spend little time on the pitch. For example, a player may come off the substitutes' bench in the final minutes of a game, score a goal with his only touch of the ball, and then never play again that season. Such a player would have an exceptionally high goal scoring rate in the wage regression, but in general may not be a high performance player. We address this issue in two ways, estimating the first-step using weighted least squares (WLS) and using robust regression as a non-parametric alternative. Using WLS, the weights for each player are proportional to the share of total minutes in the previous MLS regular season, before aggregating at the team level, thus generating alternative measures of team predicted and residual wages. We do this and summarise in Online Appendix Table C4 the second-step estimates which are comparable with our main results. Compared with Table 4 column III, there is a significant response of points per game to a team's residual wages at the 5% level, with an elasticity of 0.12. However, the effect of predicted wages is still larger and significant, with an elasticity of 0.20. With regards to home attendance, there remains no impact from team spending on predicted wages, and the impact from residual wages is significantly positive at the 5% level, with a comparable magnitude to the main results. Overall, WLS in the first step improves the precision of the second-step estimates, suggesting that this approach generates less noisy estimates of a player's or team's residual wage amount.

We also use a robust regression to reduce the weight on outliers in the estimation of Equation (3) in a non-parametric way (see e.g. Berk, 1990, for details).¹⁵ Reassuringly, we find that residual wages calculated in this manner still have a positive impact on attendance, with an estimated elasticity of 0.10, which is significant at the 10% level and quantitatively similar to our previous results (Online Appendix Table C5, column 3). As before, the highest residuals drive the results (Online Appendix Table C5, column 4). Using robust regression in the first step does, however, change the estimated effect of both the predicted and residual parts of wages on

¹⁵Specifically, we use Stata's *rreg* command with the default setting of a tuning constant equal to 7, meaning that residuals which exceed 7 times the median deviation receive zero weight, while other residuals receive a weight which decreases with their distance from the regression line.

points per game and the probability of making the playoffs: the coefficient estimates are smaller and no longer significant compared with Table 4, suggesting that the robust regression performs worse at picking up the on-pitch talent of the players from their wages. A possible reason for this is that the highest performers, due to their very high wages, receive less weight in the robust regression and, therefore, their predicted wages fall in such a way that their importance to the team is no longer captured. Indeed, we do find that the robust regression weights are decreasing with predicted wages, and that the predicted wages of the high performers are shifted down in the robust regression.

Players at the ‘bottom’ of the roster

An MLS team can comprise up to 30 players. At the bottom of a team’s roster are a number of players who may not play at all during a season, and who often earn the minimum wage specified in the CBA. To ensure that our results are not influenced by these players, we perform two final robustness checks.

First, to check whether our results are determined by ‘benchwarmers’ who play few minutes, we give zero weight in the team wage aggregates to players who played for less than 20% of the total time that their team spent on the pitch. As in our main specification in Table 4, the coefficient of a team’s aggregate predicted wages on performance is positive and significant (Online Appendix Table C6, columns I and II). However, the effect of residual wages is also positive and significant. Using home attendance as the dependent variable, the coefficient of residual wages remains significant, providing further confidence that higher spending on ‘superstar’ players translates into higher home attendances (Online Appendix Table C6, columns III and IV).

Second, there is a group of players earning the minimum wage solely determined by MLS salary regulations. The relationship between their wages and lagged performances is less clear. We perform a robustness check whereby we exclude these players from the first-step regression. As nearly 20% of our sample are paid at, or below, the senior minimum wage, excluding these players from the aggregate team wage bills in the second step is problematic.¹⁶ We therefore impute these players’ predicted wages, based on their performance in the previous season, and calculate their implied residual wages. We use these results to calculate new measures of each team’s aggregate predicted and residual wage bills. Online Appendix Table C7 shows the second-step estimates using this method. Compared with the main results (Table 4), the evidence that predicted wages affected team performance is weaker. However, the coefficient of a team’s aggregate predicted wages is significant at the 10% level when the dependent variable is whether a team reached the playoffs. As with the main results (Table 5), the estimated elasticity of home attendance to residual wages is statistically significant and driven by the 10% of players with the highest residual wages.

¹⁶This includes players aged under 24 on a team’s reserve roster, who have a lower minimum wage.

6 Conclusion

It has generally proven difficult for economists to answer why some individuals attract astronomically greater wages than their peers. We have used the market for football players in the US to investigate this phenomenon. As we were able to link player wages with measures of performance, or productivity, we could construct a measure of each player's 'predicted' wage, the part of his salary that was explained by his past performance, and his 'residual' salary. We aggregated these measures for each team, to generate their overall spending on predicted and residual wages in each season.

We found that a high spend by a team on predicted wages led to a better performance in the League, while a high spend on residual wages did not. We also found that a high spend on residual wages increased attendance at home games (a proxy for revenue), while a high spend on predicted wages did not. This latter result was driven by the players who earned the highest residual wages, above the 80th percentile. These results suggest that some players were paid large amounts because of their popularity, rather than because they were supremely talented. This is consistent with [Adler's](#) theory of superstar wages. Our results are also broadly consistent with those of [Bryson et al. \(2014\)](#), who used a similar two-step empirical strategy to investigate whether a particular sub-group of players, migrant players in Italian football, could be thought of as superstars. In addition, our results provide a possible explanation for the low model fit typically found in wage regressions within sports labour markets. We performed a number of robustness checks, exploiting the fact that our data cover almost all the players in MLS, and our main conclusions are robust to all these different model specifications.

Despite being robust, we must apply some caveats on why our conclusions may not be widely applicable. First, MLS has complicated rules regarding player salaries that are probably unique to this market. Second, a footballer's career is short, and they may only spend a few years in MLS before or after playing in other leagues. Third, MLS is still relatively new, and is different to other long-established sports leagues, particularly those in European football. Nonetheless, some features of MLS make it a useful setting to investigate the questions surrounding superstar wages. For example, it is especially competitive when compared with most other major sports leagues, such that a team's decisions about wages potentially affect its results more immediately. Likewise, MLS teams do not yet have a significant unmet demand for matchday tickets, unlike the top teams in Europe, where the majority of the global football superstars play.

Our results suggest that MLS consumers prefer to watch popular 'superstar' players. MLS competes for these players with other football leagues. It competes for consumers both with other football leagues (such as the UK's Premier League) and other sports leagues in the US (such as the National Basketball Association League). Despite this, the League's salary regulation limits the amounts that teams can spend on superstar players. It is likely, therefore, that MLS could attract more viewers by loosening this regulation and thus attracting more superstars. MLS recognised this in 2007, when it introduced the designated player rule. However, the aim of the

salary regulation is to ensure that the League remains competitive, i.e., that a small number of well-funded teams are not able to spend vastly more on actual talent than other teams and thus dominate. If MLS consumers also value a competitive league, then the salary regulation may benefit the League as a whole. Quantifying which of these effects dominates is beyond the scope of this paper, but it is an interesting area for future research.

References

- Adam, S., R. Joyce, and X. Xu.** 2019. "Labour's proposed income tax rises for high-income individuals." IFS Briefing Note BN265, Institution for Fiscal Studies.
- Adler, M.** 1985. "Stardom and Talent." *American Economic Review*, 75(1): 208–212.
- Alvaredo, F., A. B. Atkinson, T. Piketty, and E. Saez.** 2013. "The Top 1 Percent in International and Historical Perspective." *Journal of Economic Perspectives*, 27(3): 3–20.
- Arellano, M., and S. Bond.** 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2): 277–297.
- Bakija, J., A. Cole, and B. Heim.** 2012. "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data." Department of Economics Working Papers 2010-22, Department of Economics, Williams College.
- Berk, R.** 1990. "A primer on robust regression." In *Modern Methods of Data Analysis*. Eds. by J. Fox, and J. S. Long, Newbury Park, CA Sage, , Chap. 7 292–324.
- Brandes, L., E. Franck, and S. Nüesch.** 2008. "Local Heroes and Superstars: An Empirical Analysis of Star Attraction in German Soccer." *Journal of Sports Economics*, 9(3): 266–286.
- Bryson, A., G. Rossi, and R. Simmons.** 2014. "The Migrant Wage Premium in Professional Football: A Superstar Effect?" *Kyklos*, 67(1): 12–28.
- Carrieri, V., F. Principe, and M. Raitano.** 2018. "What makes you 'super-rich'? New evidence from an analysis of football players' wages." *Oxford Economic Papers*, 70(4): 950–973.
- Célérier, C., and B. Vallée.** 2019. "Returns to Talent and the Finance Wage Premium." *The Review of Financial Studies*, 32(10): 4005–4040.
- Coates, D., B. Frick, and T. Jewell.** 2016. "Superstar Salaries and Soccer Success: The Impact of Designated Players in Major League Soccer." *Journal of Sports Economics*, 17(7): 716–735.
- Ferrari, L., and M. Rueda.** 2015. "2015 MLS collective bargaining agreement: Moving in the Right Direction." *Global Sports Law and Taxation*, 6(2): 6–11.
- Franck, E., and S. Nüesch.** 2012. "Talent And/Or Popularity: What Does It Take To Be A Superstar?" *Economic Inquiry*, 50(1): 202–216.
- Franzini, M., E. Granaglia, and M. Raitano.** 2016. *Extreme Inequalities in Contemporary Capitalism*. Cham Springer.
- Frick, B.** 2007. "The Football Players' Labor Market: Empirical Evidence from the Major European Leagues." *Scottish Journal of Political Economy*, 54(3): 422–446.
- Hamlen Jr, W. A.** 1991. "Superstardom in Popular Music: Empirical Evidence." *Review of Economics and Statistics* 729–733.
- Hausman, J. A., and G. K. Leonard.** 1997. "Superstars in the National Basketball Association: Economic Value and Policy." *Journal of Labor Economics*, 15(4): 586–624.
- Hoffman, K., and C. Opitz.** 2019. "Talent and publicity as determinants of superstar incomes: empirical evidence from the motion picture industry." *Applied Economics*, 51(13): 1383–1395.

- Jane, W.-J.** 2016. “The Effect of Star Quality on Attendance Demand: The Case of the National Basketball Association.” *Journal of Sports Economics*, 17(4): 396–417.
- Jewell, R. T.** 2017. “The Effect of Marquee Players on Sports Demand: The Case of U.S. Major League Soccer.” *Journal of Sports Economics*, 18(3): 239–252.
- Kaplan, S. N., and J. Rauh.** 2013. “It’s the Market: The Broad-Based Rise in the Return to Top Talent.” *Journal of Economic Perspectives*, 27(3): 35–56.
- Koenig, F.** 2019. “Technical Change and Superstar Effects: Evidence from the Roll-Out of Television.” Working Paper.
- Krueger, A.** 2005. “The Economics of Real Superstars: The Market for Rock Concerts in the Material World.” *Journal of Labor Economics*, 23(1): 1–30.
- Kuethe, T., and M. Motamed.** 2010. “Returns to Stardom: Evidence From U.S. Major League Soccer.” *Journal of Sports Economics*, 11(5): 567–579.
- Lawson, R. A., K. Sheehan, and E. F. Stephenson.** 2008. “Vend It Like Beckham: David Beckham’s Effect on MLS Ticket Sales.” *International Journal of Sport Finance*, 3(4): 189–195.
- Lucifora, C., and R. Simmons.** 2003. “Superstar Effects in Sport: Evidence From Italian Soccer.” *Journal of Sports Economics*, 4(1): 35–55.
- Major League Soccer.** 2017a. “General Allocation Money.” , Accessed: 14/09/2020.
- Major League Soccer.** 2017b. “Targeted Allocation Money.” , Accessed: 14/09/2020.
- Major League Soccer.** 2020. “Media Resources: League Rules and Guidelines.” , Accessed: 01/02/2020.
- Major League Soccer Players Union.** 2015. “Collective Bargaining Agreement between Major League Soccer and Major League Soccer Players Union.” <https://mlsplayers.org/resources/cba>, Accessed: 01/06/2019.
- Ormiston, R.** 2014. “Attendance Effects of Star Pitchers in Major League Baseball.” *Journal of Sports Economics*, 15(4): 338–364.
- Paton, D., and A. Cooke.** 2005. “Attendance at County Cricket.” *Journal of Sports Economics*, 6(1): 24–45.
- Peeters, T.** 2015. “Profit-Maximizing Gate Revenue Sharing in Sports Leagues.” *Economic Inquiry*, 53(2): 1275–1291.
- Rosen, S.** 1981. “The Economics of Superstars.” *American Economic Review*, 71(5): 845–858.
- Simmons, R.** 1997. “Implications of the Bosman ruling for football transfer markets.” *Economic Affairs*, 17(3): 13–18.
- Smith, C.** 2013. “Major League Soccer’s Most Valuable Teams 2013.” , Accessed: 01/02/2020.
- Smith, C.** 2015. “Major League Soccer’s Most Valuable Teams 2015.” , Accessed: 01/02/2020.
- Smith, C.** 2016. “Major League Soccer’s Most Valuable Teams 2016.” , Accessed: 01/02/2020.
- Smith, C.** 2017. “Major League Soccer’s Most Valuable Teams 2017.” , Accessed: 01/02/2020.

- Smith, C.** 2018. "Major League Soccer's Most Valuable Teams 2018." , Accessed: 01/02/2020.
- Szymanski, S.** 2007. "The Economics of Sport: An International Perspective. By Robert Sandy, Peter J. Sloane and Mark S. Rosentraub." *Economica*, 74(294): 376–378.
- Terviö, M.** 2009. "Superstars and Mediocrities: Market Failure in the Discovery of Talent." *The Review of Economic Studies*, 76(2): 829–850.
- Thrane, C.** 2019. "Performance and Actual Pay in Norwegian Soccer." *Journal of Sports Economics*, 20(8): 1051–1065.
- Twomey, J., and J. Monks.** 2011. "Monopsony and Salary Suppression: The Case of Major League Soccer in the United States." *The American Economist*, 56(1): 20–28.
- US Court of Appeals.** 2002. "Fraser v. Major League Soccer." , Accessed: 01/02/2020.

Extreme wages, performance and superstars in a market for footballers

Online Appendix

Rachel Scarfe[†] Carl Singleton[‡] Paul Telemo[§]

November 2020

[†]School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh, EH8 9JT, UK;
rachel.scarfe@ed.ac.uk

[‡]Department of Economics, University of Reading, Edith Morley Building, Reading, RG6 6UB, UK;
c.a.singleton@reading.ac.uk; with thanks to the Economic and Social Research Council (UK) for funding support
under Grant No. ES/J500136/1

[§]School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh, EH8 9JT, UK;
paul.telemo@ed.ac.uk

Appendix A. Further details on data and variables

TABLE A1: Notes on variables

Variable name	Notes/Description	Source
<i>Player-level variables:</i>		
Position	Derived primary position as of MLS records in 2018. The positions are forward, midfielder, defender (outfield players) or goalkeeper; some players may have played in multiple positions throughout their MLS careers.	MLS
Multi-position	Dummy variable for a player who can play in more than one primary position	MLS
<i>Player-season-level characteristic variables:</i>		
Wages/salary (w)	Log guaranteed annual salary (US\$)	MLSPA
Age	Player age in years (minus 15) at the beginning of each season	MLS
Captain	Dummy variable for the team captain	Various, MLS website
Designated	Dummy variable for a designated player	MLS
MLS new entrant	Dummy variable for an entrant to MLS, who had not played in the MLS before (during the sample period)	Derived
MLS re-entrant	Dummy variable for an entrant to MLS who had previously played in MLS during the sample period but who was not signed to a team in the previous season	Derived
MLS transfer	Dummy variable indicating a player's first season at a new team, after playing for another MLS team the previous season.	Derived
Mins played (%)	Percentage of the team's regular season minutes played (i.e. time on the football pitch)	MLS
Rating	A combination score of a player's average performance in the season	WhoScored.com via Opta
Page views	Average of views of a player's Wikipedia profile page in January and February	Wikipedia API
<i>Player-season-level performance variables (all derived as the average per 90 minutes played):</i>		
Goals	Goals scored (outfield players only)	MLS
Assists	Number of passes to another player that result in a goal being scored (outfield players only) - a higher value would indicate a more productive player, especially for forwards and midfielders	MLS

(Table A1 continued)

Variable name	Notes/Description	Source
<i>Player-season-level performance variables (given as the average per 90 minutes played)</i>		
Shots	Number of shots toward goal, both on and off target - a higher value would indicate a more productive player, especially for forwards and midfielders	MLS
Shots on goal	Number of shots on target that could have resulted in a goal, including those that were saved by a goalkeeper or blocked by another player - a higher value would indicate a more productive player, especially for forwards and midfielders	MLS
Fouls committed	Number of fouls committed by the player, including minor fouls. A higher value may suggest a player prepared to take more risks - a priori it is not obvious whether this is indicative of a higher productivity player	MLS
Fouls suffered	Number of times the player suffered a foul committed by another player - higher value suggests a player may be adept at attracting fouls and may indicate higher productivity, though conversely it may indicate a lack of assertiveness by players	MLS
Red cards	Number of red cards for serious fouls, that result in the player being sent off the pitch. A higher value may suggest a player prepared to take more risks - a priori it is not obvious whether this is indicative of a higher productivity player	MLS
Saves	Saves as a percentage of shots on goal by opposing players (goalkeepers only) - a higher value would indicate a more productive player	MLS
<i>Team-season-level variables:</i>		
Log points per game	Natural logarithm of the total points achieved over the regular season divided by the number of games played	MLS
Log attendance	Log of average attendance (10,000 persons) at home games during the season	MLS
Playoffs	Dummy variable indicating whether a team qualified for the MLS playoffs due to their performance during the regular season	MLS
Expansion team	Dummy variable for a team's first season in MLS	Derived
Log revenue	Log of estimated team revenue (US\$, millions), available for 2013, 2015-18 seasons	Forbes
Log value	Log of estimated team value (US\$, millions), available for 2013, 2015-18 seasons	Forbes

TABLE A2: Number of players by year and position in analysis sample

Year	Defender	Forward	Goalkeeper	Midfielder	Total
2008	126	80	43	146	395
2009	123	75	40	127	365
2010	118	82	39	135	374
2011	142	113	53	194	502
2012	142	123	58	202	525
2013	148	130	59	197	534
2014	161	127	55	187	530
2015	149	119	58	206	532
2016	157	99	60	197	513
2017	184	108	61	213	566
2018	206	131	74	211	622
Total	1656	1187	600	2015	5458

Notes.- player positions are defined according to the primary record as observed on the MLS website in 2018.

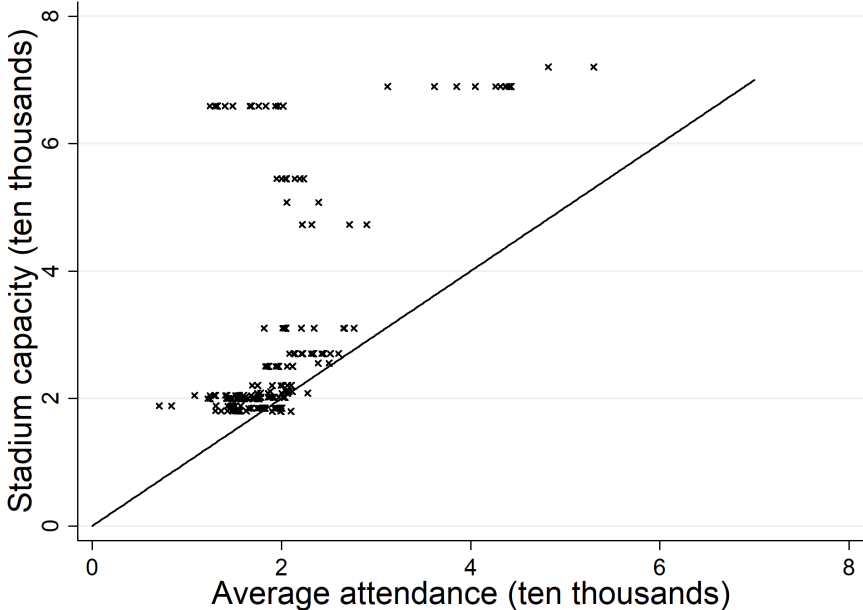
TABLE A3: MLS teams/franchises & abbreviations, 2007-18

Abbreviation	Latest team name	Years active in sample period
ATL	Atlanta United FC	2017-18
CHI	Chicago Fire	2007-18
CHV	Chivas USA	2007-14
CLB	Columbus Crew	2007-18
COL	Colorado Rapids	2007-18
DAL	FC Dallas	2007-18
DC	D.C. United	2007-18
HOU	Houston Dynamo	2007-18
KC	Sporting Kansas City	2007-18
LA	LA Galaxy	2007-18
LAFC*	Los Angeles FC	2018
MIN	Minnesota United FC	2017-18
MTL	Montreal Impact	2012-18
NE	New England Revolution	2007-18
NYCFC	New York City FC	2015-18
NYRB	New York Red Bulls	2007-18
ORL	Orlando City SC	2015-18
PHI	Philadelphia Union	2010-18
POR	Portland Timbers	2011-18
RSL	Real Salt Lake	2007-18
SEA	Seattle Sounders FC	2009-18
SJ	San Jose Earthquakes	2007-18
TOR	Toronto FC	2007-18
VAN	Vancouver Whitecaps FC	2011-18

Notes.- *LAFC joined MLS in 2018, the last year of our sample period. At the time of writing, team-level data was not available for LAFC in 2018, so we do not include it in the second step of our analysis. We do include its players in the first step of our analysis.

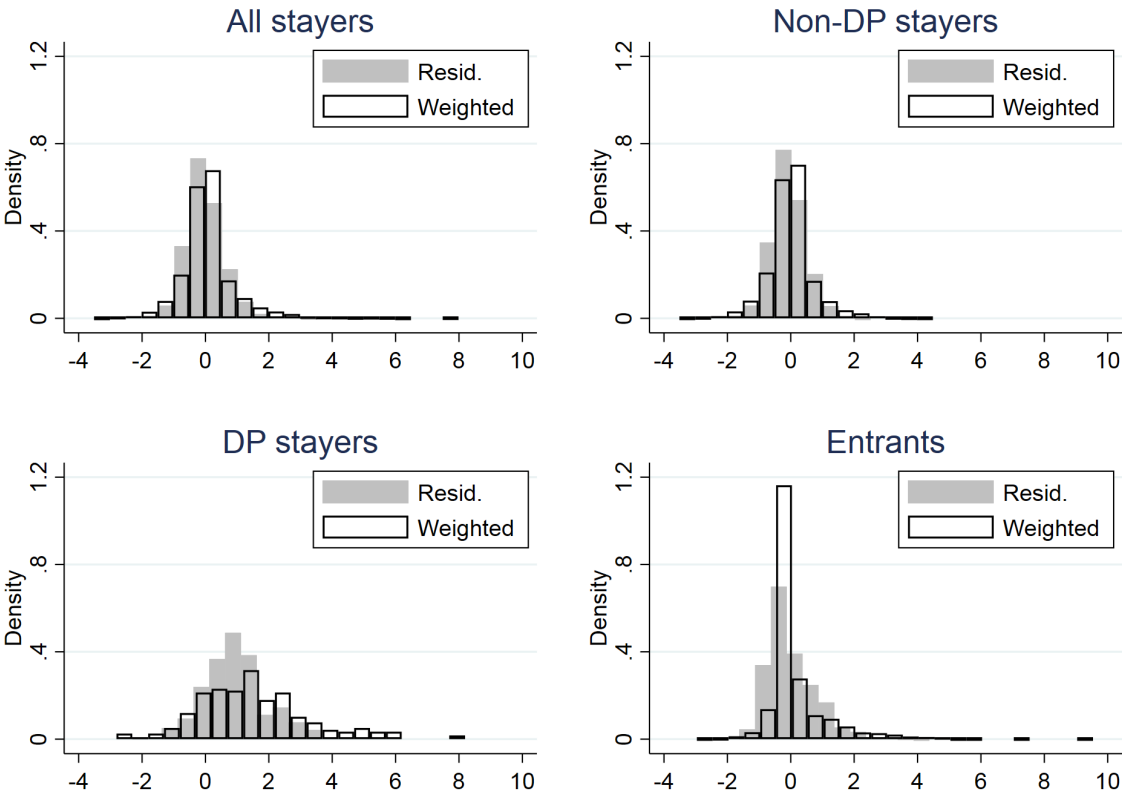
Appendix B. Additional figures

FIGURE B1: Average home attendance vs stadium capacity, by team-season, 2007-18



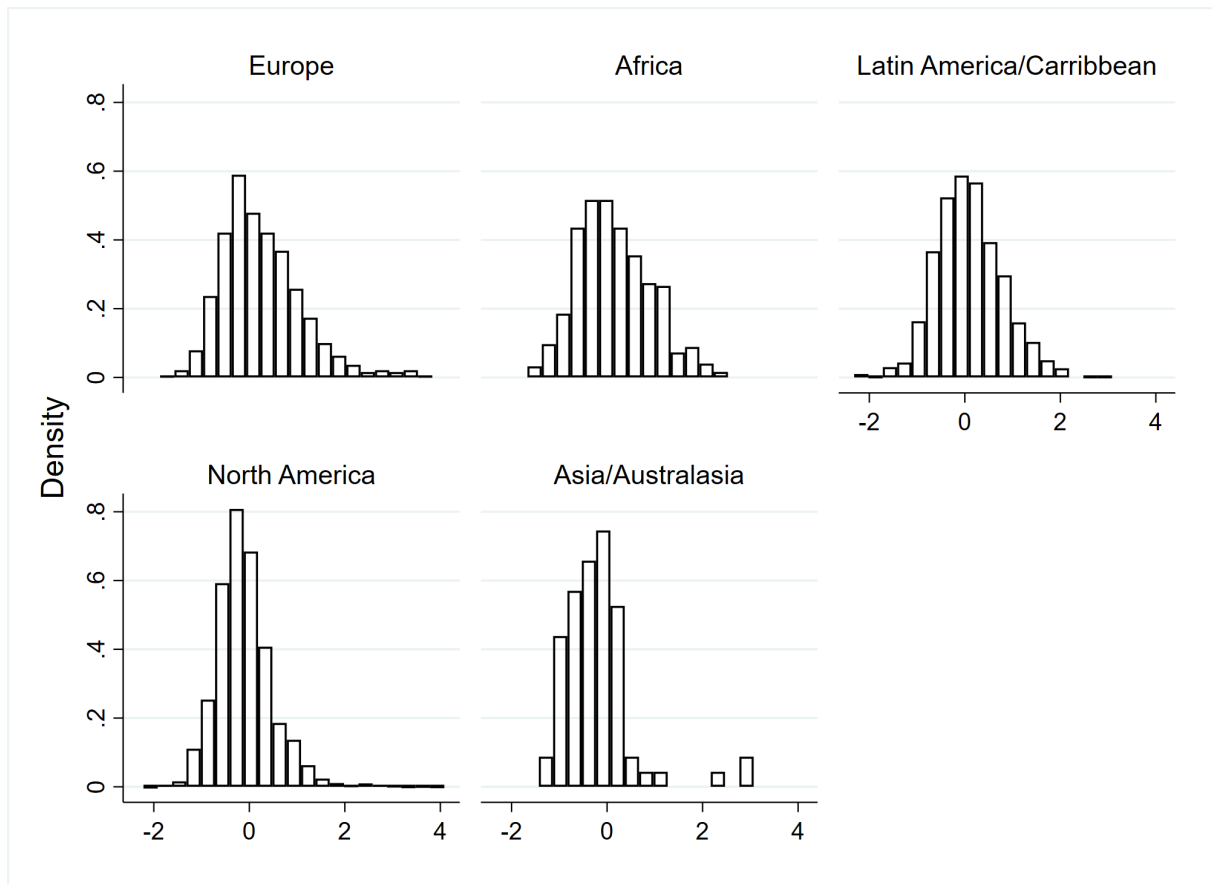
Notes.- author calculations using MLS and other sources. It is possible to be below the 45-degree line; many teams have the potential to increase their stadium capacity within a season, by agreeing with stadium owners for one-off matches to sell tickets in parts of the stadia that are normally not used.

FIGURE B2: Distributions of estimated residuals from first-step player salary regressions, stayers vs entrants to MLS and unweighted vs weighted models, 2008-18



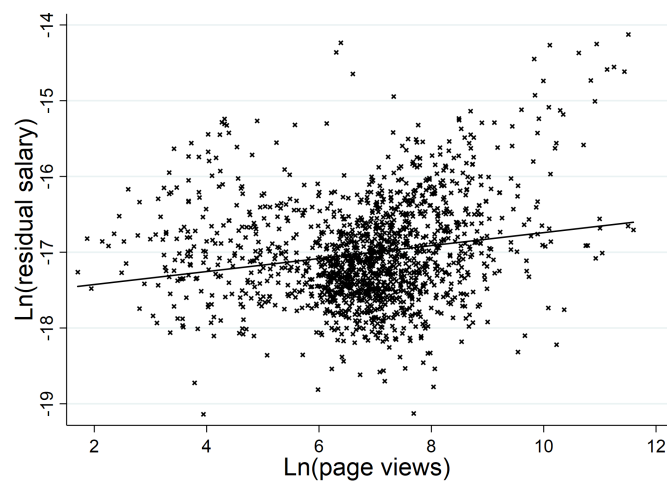
Notes.- See Figure 4. A ‘stayer’ is a player-year observation who was in MLS the preceding season. An ‘entrant’ is a player-year observation who was not in MLS the preceding season, either because they have entered for the first time or have returned to the league. ‘DP’ refers to players with designated player status.

FIGURE B3: Distributions of estimated residuals from first-step player salary regressions, by player region, 2008-18.



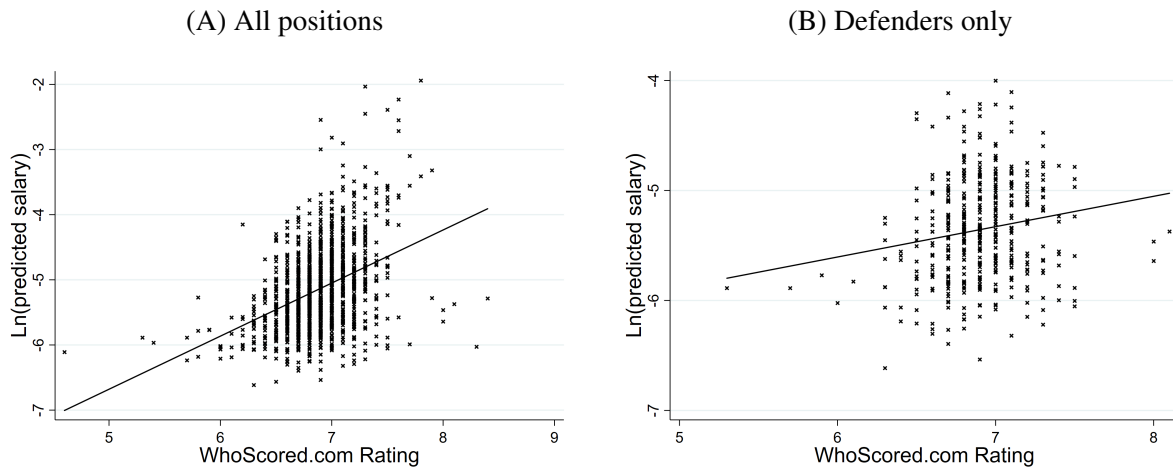
Notes.- See Figure 4.

FIGURE B4: Correlation between the residual salaries and Wikipedia page views of players in MLS, 2016-18



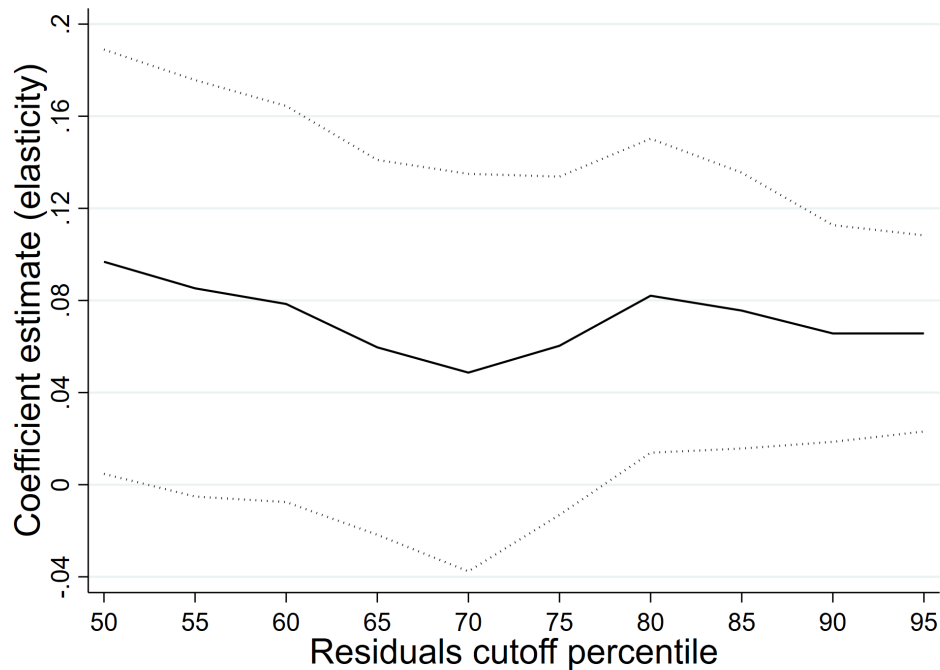
Notes.- solid line plots the line of best fit. Residual salary is measured for each player-year observation from estimates of Equation (3). Wikipedia page views are only for the off-season months of January and February.

FIGURE B5: Correlation between the predicted salaries and previous season whoscored.com rating of players in MLS, 2014-18



Notes.- solid lines plot the lines of best fit. Predicted salary is measured for each player-year observation from estimates of Equation (3). Shows the subsample of 1,299 player-year observations (428 defenders) in MLS seasons 2014-18 who had a [Whoscored.com](#) rating in the previous season.

FIGURE B6: Sensitivity analysis of $\hat{\gamma}_H$: varying the definition of 'high' residual salary



Notes.- the solid line plots the estimated effects of aggregate residual salary above the Xth percentile on team average home attendance, i.e., varying the cutoff, equivalent to the results presented in column IV of Table 5, which used the 90th percentile (also shown here). Dotted lines show 95% confidence intervals robust to team-level clusters.

Appendix C. Additional tables

TABLE C1: Robustness check: regression of log residual salary on players' Wikipedia page views

	(I)	(II)	(III)
Ln(page views)	0.086*** (0.020)	0.085*** (0.020)	0.084*** (0.020)
Constant	-17.60*** (0.135)		
Year fixed effects	No	Yes	Yes
Team fixed effects	No	No	Yes
R^2	0.036	0.036	0.048
N	1,474	1,474	1,474

***,**, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to player clusters and displayed in parentheses. See also Appendix Figure B4.

TABLE C2: Robustness check: regression of log predicted salary on previous season's WhoScored.com player rating, 2014-2018

	(I)	(II)	(III)	(IV)
Rating	0.815*** (0.094)	0.847*** (0.094)	0.849*** (0.090)	0.348*** (0.092)
Constant	-10.80*** (0.631)			
Year fixed effects	No	Yes	Yes	Yes
Team fixed effects	No	No	Yes	Yes
R^2	0.159	0.170	0.197	0.105
N	1,299	1,299	1,299	429

***,**, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to player clusters and displayed in parentheses. Column (IV) is only for players whose main position is defender. See also Appendix Figure B5.

TABLE C3: Robustness check, second-step regression estimates: log annual revenue and value (Forbes, \$millions), 2013, 2015, 2017 and 2018

	Revenue		Value	
	(I)	(II)	(III)	(IV)
Log points per game	0.061 (0.135)		0.112 (0.100)	
Log attendance (10,000s) ($\hat{\phi}_2$)	0.677*** (0.181)	0.694*** (0.192)	0.666*** (0.198)	0.633*** (0.166)
<i>Log wages, all players (\$m):</i>				
Predicted ($\hat{\lambda}$)	0.564* (0.280)	0.558* (0.276)	0.225 (0.204)	0.216 (0.201)
Residual ($\hat{\gamma}$)	0.055 (0.082)		-0.007 (0.062)	
Residual above 90th percentile ($\hat{\gamma}_H$)		0.009 (0.029)		0.008 (0.021)
Year fixed effects	Yes	Yes	Yes	Yes
Team fixed effects	Yes	Yes	Yes	Yes
N	95	94	95	94
R^2	0.845	0.841	0.949	0.948

***, **, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters (20) and displayed in parentheses.

TABLE C4: Robustness check, second-step regression estimates: weighted least squares in first step, 2008-18

	Points per game	Playoffs	Attendance	
	(I)	(II)	(III)	(IV)
Log points per game			0.147** (0.069)	0.132** (0.065)
Expansion			0.082 (0.052)	0.079 (0.051)
<i>Log salary (\$m):</i>				
Predicted			0.015 (0.057)	0.004 (0.060)
Residual			0.122*** (0.043)	
Bottom 90% residual				0.102 (0.091)
Top 10% residual				0.051*** (0.018)
<i>Log weighted salary (\$m):</i>				
Predicted	0.199*** (0.069)	1.853*** (0.498)		
Residual	0.117** (0.057)	0.854* (0.475)		
Year fixed effects	No	No	Yes	Yes
Team fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	204	186	204	199
<i>R</i> ²	0.284		0.858	0.865

***, **, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters (19 for Column (II), 23 otherwise) and displayed in parentheses.

Columns here are the equivalents of Table 4:(III), Table 4:(VI), Table 5:(I) and Table 5:(IV), respectively. First step regression models estimated using the team-season percent of minutes played for each observation as weights.

TABLE C5: Robustness check, second-step regression estimates: robust regression in the first step, 2008-18.

	Points per game	Playoffs	Attendance	
	(I)	(II)	(III)	(IV)
Log points per game			0.137** (0.068)	0.107 (0.065)
Expansion			0.105** (0.049)	0.095* (0.048)
<i>Log salary (\$m):</i>				
Predicted			0.125 (0.103)	0.071 (0.094)
Residual			0.101** (0.049)	
Bottom 90% residual				0.075 (0.123)
Top 10% residual				0.051** (0.021)
<i>Log weighted salary (\$m):</i>				
Predicted	0.055 (0.054)	0.447 (0.307)		
Residual	0.052 (0.056)	0.439 (0.366)		
Year fixed effects	No	No	Yes	Yes
Team fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	204	186	204	196
<i>R</i> ²	0.254		0.860	0.889

***, **, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters (19 for Column (II), 23 otherwise) and displayed in parentheses.

Columns here are the equivalents of Table 4:(III), Table 4:(VI), Table 5:(I) and Table 5:(IV), respectively. First step regression models estimated using robust regression.

TABLE C6: Robustness check, second-step regression estimates: assigning players who played less than 20% of total time a weight of zero, 2008-18.

	Points per game	Playoffs	Attendance	
	(I)	(II)	(III)	(IV)
Log points per game			0.146** (0.068)	0.135** (0.061)
Expansion			0.082 (0.052)	0.079 (0.051)
<i>Log salary (\$m):</i>				
Predicted			0.009 (0.059)	-0.003 (0.063)
Residual			0.127*** (0.045)	
Bottom 90% residual				0.112 (0.090)
Top 10% residual				0.061*** (0.020)
<i>Log weighted salary (\$m):</i>				
Predicted	0.212*** (0.068)	1.840*** (0.511)		
Residual	0.116** (0.054)	0.867* (0.464)		
Year fixed effects	No	No	Yes	Yes
Team fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	204	186	204	197
<i>R</i> ²	0.288		0.858	0.868

***, **, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters (19 for Column (II), 23 otherwise) and displayed in parentheses.

Columns here are the equivalents of Table 4:(III), Table 4:(VI), Table 5:(I) and Table 5:(IV), respectively. First step regression models estimated using the team-season percent of minutes played for each observation as weights.

TABLE C7: Robustness check, second-step regression estimates: excluding minimum wage players from first-step estimation, 2008-18.

	Points per game	Playoffs	Attendance	
	(I)	(II)	(III)	(IV)
Log points per game			0.143** (0.063)	0.144** (0.056)
Expansion			0.085 (0.052)	0.086 (0.054)
<i>Log salary (\$m):</i>				
Predicted			0.035 (0.178)	0.012 (0.138)
Residual			0.176*** (0.060)	
Bottom 90% residual				0.188 (0.142)
Top 10% residual				0.071*** (0.022)
<i>Log weighted salary (\$m):</i>				
Predicted	0.193 (0.166)	1.518 (1.019)		
Residual	0.124 (0.091)	1.000* (0.548)		
Year fixed effects	No	No	Yes	Yes
Team fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	204	186	204	196
<i>R</i> ²	0.265		0.862	0.872

***, **, * indicate significance from zero at 1%, 5% and 10% levels, respectively, two-sided tests, standard errors robust to team clusters (19 for Column (II), 23 otherwise) and displayed in parentheses.

Columns here are the equivalents of Table 4:(III), Table 4:(VI), Table 5:(I) and Table 5:(IV), respectively.