Are Soccer Teams Being Inefficient? An Analysis of Sunk Cost Fallacy and Recency Bias Using Transfer Fee

Junru Lyu

December 2020

Department of Economics

University of California, Berkeley

Undergraduate Honors Thesis

Advised by Professor Stefano DellaVigna

Abstract

Sunk cost fallacy and recency bias are frequently discussed phenomenons in behavioral economics. While there is experimental evidence demonstrating the existence of both phenomenons, relevant field studies are limited, and researchers have even come to distinct conclusions using empirical analysis in sports labor markets. Hence, this paper adds to the existing literature and attempts to detect sunk cost fallacy and recency bias using field evidence among soccer teams in the Top Five Soccer Leagues using panel data of soccer transfers from 2000 to 2020 with performance control. For sunk cost fallacy, the paper tries to find out whether teams give more playing minutes to players that they paid more for, holding performance constant. The results show that the correlations between playing time and transfer fee are statistically significant within the first three years after transfers occur, suggesting presence of sunk cost fallacy that decreases in economic magnitude as time persists. The paper also deals with unobserved variable bias and tests for coefficient stability using methods developed by Altonji et al. (2000) and Oster (2017). For recency bias, the paper tests whether teams, when setting transfer fees, pay too little attention to player performance in earlier years compared to the relative importance of the earlier performance in determining performance after transfer. Using the difference between ratios of coefficients of past performance prior to transfer on transfer fee and on performance after transfer, the paper finds evidence of soccer teams committing to recency bias. The standard errors generated by bootstrapping, however, suggest that result is statistically insignificant.

Acknowledgement

I thank Professor Barry Eichengreen, Matthew Tauzer, and Javier Feinmann for their suggestions. I especially appreciate my advisor, Professor Stefano DellaVigna, for his guidance and continued support throughout the composition of this paper. I take full responsibility for any shortcomings remaining.

1 Introduction

Proposed by Thaler (1980), sunk cost fallacy is a frequently discussed behavioral phenomenon where people are willing to endure additional, and likely unnecessary, cost for prior investment. A classic example is, after people pay for movie tickets, they are often willing to suffer through an unenjoyable movie that they otherwise prefer not to watch free of charge. While we, in everyday life, may commit to sunk cost fallacy for only a few dollars, soccer teams often face situations where the clubs have invested millions or even tens of millions of euros on soccer player transfers without obtaining expected returns.

In soccer, player transfer is a process where a club obtains a desired player through a variety of methods. The most well-known method involves a soccer club paying another club an amount of monetary value to buy the player out of his contract and sign the player. Other methods also involve free transfers, where a player moves to another team after his contract expires, and signing from the youth teams, which is also free. The team is, then, willing to pay for a certain amount for a player and it expects the player to prove his worth of value by his output on the field. However, it is possible that a soccer team may commit to sunk cost fallacy and tend to give player with higher transfer fee more time to play even though the player's performance may not be worthy of the time given.

With panel data from 2000 to 2020 of transfer fee, playing time, and performance, this research paper explores whether soccer teams fall into the trap of sunk cost fallacy. Using ordinary least square (OLS), the paper finds positive correlation between transfer fee, which is the sunk cost, and playing time. However, there is likely endogeneity issues between transfer fee and playing time as players with higher transfer fee are likely more skilled, which encourages teams to provide more playing time to them. As many variables that determine players' ability are unobservable, the endogeneity issue causes the regression with only observed variables to be biased upward. Hence, the paper also attempts to deal with unobserved variable bias by employing methods developed by Altonji et al. (2000) and Oster (2017) to examine coefficient stability in the presence of unobserved factors. Specifically, the paper studies how the coefficient on transfer fee varies as more unobserved controls are added by extrapolating the change in coefficient with gradually added control variables. Evidence of sunk cost persists after accounting for unobserved variable bias. Unfortunately, additional tests for causation of transfer fee on playing time, using instrumental variable regression with instruments being balance of the team selling the players, fail to deliver significant result, so there is insufficient evidence to account higher payment as the causation of more playing time.

Another phenomenon in behavioral economics that may be present among soccer teams is recency bias, introduced by Tversky and Kahneman (1973). It occurs when people mostly remember things that occur recently and tend to ignore occurrences in the past. Recency bias, then, may cause teams to evaluate a player primarily based on his performance in the immediate season before transfer. Indeed, as soccer players' performance change frequently and, sometimes, quite intensely in short period of time, soccer teams may focus mainly on recent performance when they decide to purchase players. Since a soccer player's future performance may largely depend on his performance in the present, soccer teams may be correct in evaluating a player's worth mainly through his recent output. However, are teams focusing too much on the recent performance and somewhat neglecting players' performance in the past? In this case, teams may be over or underestimating players' values. The paper, then, regresses transfer fee and performance after transfer on performance prior to transfer to detect whether more recent performance may affect transfer fee more than it affects performance after transfer. In particular, the paper compares ratios of coefficients of past performance prior to transfer on transfer fee and on performance after transfer to show that soccer teams indeed commit to recency bias by overestimating the impact of more recent performance when paying transfer fee. The standard errors of the ratios generated by bootstrapping, however, suggest that the result found is statistically insignificant.

If team performance is considered as the revenue, the finding of sunk cost fallacy and recency bias may be extended to discuss whether these phenonmenons are causing firms to not maximize their profits and to become inefficient and thus irrational, an issue that concerns companies in most if not all industries and is thus worth to be continuously explored. This paper, then, adds to existing literatures relevant of behavioral economics phenomenons using data from sports industry and demonstrates that data from soccer, which is relevant to many fields of studies in economics and in social sciences, is capable of producing meaningful economic results.

For the rest of the paper, section 2 gives a review of existing literature relevant to the paper, section 3 describes the data and empirical methods, section 4 presents the results, section 5 discusses the interpretations of results and any shortcomings, and section 6 concludes the paper.

2 Literature Review

Thaler (1980) introduces sunk cost fallacy by constructing a list of irrational behaviours that are inconsistent with traditional economic theory but are crucial to the foundations of behavioral economics. It is an irrational phenomenon where people are willing to endure irrational cost for prior investment (Thaler, 1980). While Thaler (1980) does not provide experimental evidence or any data associated with sunk cost fallacy, the idea provides the intuition and foundation of many other research topics. To experimentally test sunk cost fallacy, Arkes and Blumber (1985) construct a series of survey to show that people indeed would fall in the trap of sunk cost fallacy. However, it remained quite difficult to perform quantitative field test for such phenomenon.

Staw and Hoang (1995) become one of the first to quantitatively test for sunk cost fallacy using National Basketball Association (NBA) data. They construct a player performance index, and, controlled on performance, they find that higher draft pick position, as a sunk cost, correlates with more playing time in the first five years after they were drafted (Staw & Hoang, 1995). This research method lays the experimental foundation of finding evidence of sunk cost fallacy using sports data. I also use playing time as the dependent variable and investigate the effect of the sunk cost, which is transfer fee, controlled on player performance. Besides the fact that our studies differ in specific industries, the setup of our analysis for sunk cost fallacy are highly similar.

Staw and Hoang (1995) have started a trend of exploring behavioral economics phenomenon in the realm of sports. As an example, Camerer and Weber (1998) revisit the findings by Staw and Hoang (1995) using additional performance controls and a trade factor (which takes into account whether players are traded after they are picked); the sunk cost fallacy persist but much less in magnitude. There are also other attempts to study sunk cost fallacy using NBA data recently. In particular, Leeds et al. (2013) use regression discontinuity on first and second round picks to demonstrate that there is in fact no observed sunk cost fallacy in NBA and argue that traditional economic theory would predict NBA coaches' choices better than behavioral theory. The set up of the paper is similar to that of Staw and Hoang (1995) but the method is drastically different. Instead of OLS, Leeds et al. (2013) use regression discontinuity between the draft rounds as an instrument and find no evidence of sunk cost fallacy. Unfortunately, such method is not applicable to sunk cost fallacy analysis using transfer fee. The only source of discontinuity that may exist is using the discontinuity between transfers with strictly positive transfer fee and free transfers. However, if free transfer is considered as the discontinuity, there is no variation with the among those who are transferred with no cost.

In addition, Hinton and Sun (2020) use panel data with fixed effects to account for unobserved variables and demonstrate that NBA player salary, as a sunk cost, has a small but statistically significant effect on playing time. The approach is different from prior results since the authors use monetary values as a sunk cost whereas all the prior papers have used draft picks. Salary information is much less transparent in European soccer leagues, which makes the approach less feasible; however, the concept of transfer fee is very similar to salary. In fact, transfer fee is a one-time payment while salary is continuous throughout the year, so, in a sense, transfer better fit into the classification of sunk cost than salary does.

The use of sports data is not limited to only Basketball. Researchers have attempted to detect sunk cost fallacy in Australian and American football[1]. Borland et al. (2011) use a similar idea as Staw and Hoang (1995) and explore the impact of draft position in the Australian Football League (AFL), controlled on several performance statistics in Australian Football, on games played and tenure status. However, the effect is much weaker than that of Staw and Hoang (1995); the paper attributes the finding to the different structures between AFL and NBA as coaches in AFL are "evaluated primarily on the basis of their ability to win matches." (Borland et al., 2011) Keefer (2015, 2017, 2019), on the other hand, has more success with finding sunk cost fallacy using abundant data from National Football League (NFL). Using a subset of defensive players, Keefer (2015, 2017) finds that higher salary cap values have significant impacts on playing time and starting frequency with both ordinary least squares (Keefer, 2015) and regression discontinuity using draft pick rounds controlled on player performance (Keefer, 2017). Even with within-game analysis, Keefer (2019) finds that running backs with higher salary cap has more rushing attempts, which indicates higher within game usage of the running back.

[1]The terminology of soccer and football are often used interchangeably, but the references of football in Australian and American football related papers are different from 'soccer.' Hence, this paper will refer to the sport 'soccer' only as 'soccer' unless otherwise specified.

Despite long tradition of using sports data to test for sunk cost fallacy, there has been limited usage of data in soccer for sunk cost fallacy or even for behavioral economics in general. While Mourao (2016) uses transfer fee to systematically calculate club spending efficiency, the application of soccer transfer fee in sunk cost fallacy, or even behavioral economics, is recent, when Hackinger (2019) tests but finds no evidence for the presence of sunk cost fallacy in the soccer industry using transfer fees. However, as Hinton and Sun (2020) mention, the control measure can often determine whether significant sunk cost fallacy is found. The difficulty in obtaining the control is that there is not yet a performance measure that evaluates soccer players across all positions. Hackinger (2019) uses seasonal players ratings from Kickers, a magazine that evaluates only the German league, to control for player performance, but seasonal ratings are likely biased toward players who have more playing time and are limited. As a result, this paper chooses to not rely on ratings and instead use by-minute statistics (such as goals and assists) as a control of performance.

Such an attempt is more difficult in soccer than in basketball and football due to great variability in position. In soccer, players of different positions have very different functions: some defenders may never rarely go into opponent's half field while some forwards may never be a part of defense. While Keefer (2015) offers the idea to limit to a strict subset of the position, the larger issue lies in that players of the same position can have very different functions. Some midfielders may be valued for their accurate passes while others for their interceptions. Even among the defenders, some are known to be the last line of defense while others may be expected to join the offense more. While goalies' main purpose is uniformly to save as many shots as possible, there are rarely rotations in this position unless for injury or suspension purposes, so there is little variation. Therefore, in soccer, forward is arguably the best choice. While there is also slight variation among forwards, they are almost always expected to create as many goals as possible either through goals or assists. Unlike goal keepers, there is also variation in forwards as there are plenty of rotations.

Still, there may be unobserved variables that we cannot observe but the teams do. Practice quality and leadership are examples of factors that may determine play time but are unobservable or not quantifiable. To address this issue, I use a method developed by Altonji et al. (2000) and Oster (2017) to analyze coefficient stability of regressions with unobserved variables. Altonji et al. (2000) introduce the idea that, if the observable variables are a random subset of all determining factors of the dependent variable, then the true effect of the independent variable on the dependent may be recovered by imposing certain assumption regarding the importance of observable variables relative to that of the unobserved variables. Altonji et al. (2000) use this method to demonstrate that Catholic schools has significant positive impact on attendance of high school and university. Oster (2017) extends this idea and proposes a method that uses the relationship between the R-squared and coefficient of univariate regression and those of regression with full set of observable controls to recover the coefficient without unobserved variable bias. This paper uses the formula derived by Oster (2017) to determine whether transfer fee has a positive correlation with playing time after such biases are eliminated. The details of the formula and implementation will be in Section 3.

On the other hand, such method will not be imposed for the recency bias analysis since the unobserved variables are unlikely to significantly impact the results. The key controls for recency bias are performance before a player transfers. In this case, the teams that purchase also cannot observe the set unobserved variables, such as practice quality and player characteristics. Hence, it is likely that teams purchasing players use the same set of observed and quantifiable variables as this paper uses to make decisions. In this case, the paper uses similar methodology and empirical set up as Healy (2008) uses. Healy (2008) employs a self-constructed measure to control for performance of players in Major League Baseball (MLB); Healy (2008) finds that teams in MLB tend to over-value recent performance as the ratio of coefficient of past performance over that of more recent performance is significantly lower on salary than on future performance. Hence, the relative importance of recent performance is much higher for salary than for future performance, which shows that baseball teams tend to commit to recency bias. This paper uses a similar set up but with transfer fee instead of salary as the other dependent variable. Moreover, Healy (2008) uses bootstrapping to determine the standard error of the ratio of coefficients, which will also be employed in this paper.

Hochberg (2011) also uses MLB data to demonstrate that teams likely overweight players' performance in the year before players' current contract performance (which is also known as contract year) expires when deciding how much salary to pay players. The ratio between coefficients of performance three years before new contract over performance of contract year regressed on salary is statistically lower than that regressed on performance after signing a new contract. Hence, Hochberg (2011) confirms the existence of recency bias in baseball.

Similar works also extend to basketball with NBA data regarding team performance. Camerer (1989) demonstrates that the betting market tends to overvalue a team's point break, which is the forecast in the betting market regarding how many points a team may win (or lose) in games, when a team is on a winning streak. Moreover, Camerer (1989) demonstrates that bettors betting for teams on losing streak and against teams on winning streaks are more likely to break even. Hence, despite that such bias is too economically small to exploit profit (Camerer, 1989), the market also commits to recency bias by overvaluing teams who perform better recently. Fox (2015) adds to study of recency bias in basketball by using player performance data in NBA to demonstrate that more recent performance is over-weighted by teams when they decide on the salary and duration of players' new contract. The ratio between coefficient of performance two years and three years before contract expiration over that of contract year on regressions with future player and team performance is significantly lower than those on total contract value, salary, and contract duration. Hence, it seems that both betting markets and teams in basketball tend to commit to recency bias.

These sources of irrationality regarding recency bias originates from the study of heuristics by Tversky and Kahneman (1973). Tversky and Kahneman (1973) argue that people sometimes commit to "judgmental heuristic in which a person evaluates the frequency of classes or the probability of events by availability." In other words, people make decisions based on "the most relevant instances" that come to mind rather than a holistic review (Tversky & Kahneman, 1973). In the case of recency bias, the relevant instances are the recent performance of players and teams, which causes teams and betting markets to heuristically judge players and teams future outcome, resulting in inefficient and irrational outcomes.

With a valid control, there is great hope in finding evidence of sunk cost fallacy and recency bias in the paper since there is more evidence finding the existence of sunk cost fallacy and recency bias than there is that disprove it. In the realm of basketball, three out of the four papers find evidence of sunk cost fallacy with draft picks and salary (Camerer & Weber, 1998; Hinton & Sun, 2020; Leeds et al., 2013; Staw & Hoang, 1995). In football, Keefer (2015, 2017, 2019) also find evidence of sunk cost fallacy with respect to multiple measures of sunk cost.

However, there is reason to believe that there may not be evidence of sunk cost fallacy in soccer due to different structures of the league. In European soccer leagues, there is relegation, meaning that teams with the worst records are dropped to second division league, so there is less room for the teams to commit to sunk cost fallacy (Hackinger 2019). Moreover, it seems that most papers who study sports in the United States (US) tend to find evidence of the existence of both sunk cost fallacy (Camerer & Weber, 1998; Hinton & Sun, 2020; Keefer, 2015, 2017, 2019; Staw & Hoang, 1995) and recency bias (Camerer, 1989; Fox, 2015; Healy, 2008; Hochberg, 2011). In fact, only Leeds et al. (2013) find no evidence of irrationality when studying a market in the US. In contrast, the only papers that study sports market outside of the US fail to detect evidence of these phenomenons (Borland et al., 2011; Hackinger, 2019). Such patterns may be due to that US and other countries evaluate coaches abilities differently (Borland et al., 2011). In the United States, there may be pressure for coaches and managers in all sports to play or sign players who are popular and are able to help the teams to generate more profits. In this case, those players who are popular are likely to have associated sunk cost (such as higher draft pick, higher salary, or better recent performance), which causes team to commit to sunk cost fallacy or recency bias. On the other hand, Borland et al. (2011) argue that coaches in other countries are mainly judged by team performance, so they are less likely to fall into the trap of sunk cost fallacy.

Yet, this paper argues that all teams face pressure to receive better records. Both NBA and NFL teams fight to be in the play-offs, so all sports teams have incentives to achieve better performance and thus little room to commit to sunk cost fallacy. Moreover, most US leagues face a salary cap, so there is even less room to commit to recency bias. On the other hand, soccer teams do not face restricted salary or transfer fee cap, so there are less restrictions to prevent them from committing sunk cost fallacy and recency bias. As a result, the paper believes that there will likely be evidence of sunk cost fallacy and recency bias in soccer teams due to similar competition environment and goals of the sports teams. The result the paper is, to its knowledge, the first to use with-in game statistics of soccer players to evaluate behavioral economics concept. If the controls are proven viable, data from the soccer industry—with its abundant performance and managerial data—may potentially open up more economics studies.

3 Data and Empirical Framework

3.1 Description of Data

This paper specifically focuses on data from the top five European soccer leagues, which are the Premier League of England, Bundesliga of Germany, La Liga of Spain, Serie A of Italy, and Ligue 1 of France. The paper selects these leagues because clubs from these competitions represent the highest level of competitiveness and professionalism in the world. In addition, due to that the competitions are all in Europe, there are many transfers within teams from these leagues and theses leagues have very similar environments in terms of competitiveness of the games and level of the amount of financial transactions, so teams from these leagues are comprehensive and comparable.

The entire dataset, which includes all transfers and player performance within the Top 5 league from 2000 to 2020, was scraped from transfermarkt.com, the same website that Hackinger (2019) used and arguably the most credible website that keeps track of player and team records. For scraping method, the paper uses the module BeautifulSoup on Python Jupyter Notebook with self-written codes. An issue arising from using monetary value as sunk cost is that same amount of money likely worth differently for different teams. Luckily, transfermarkt provided enough data to implement fixed effects for teams and positions, which allows the analysis to compare transfer fee of players of a similar position on the same team.

For the main variables, transfer fee is in euros, which I first adjust for inflation then log to reduce variation, and total playing time measures minutes played in a season. As only forward players are used for analysis, the best performance indices are goal and assist efficiency, which are constructed using goals and assists divided by minutes played. Points per appearance measures the number of points the team may earn every time a player plays, so this variable captures the amount a player contributes to team performance; hence, it is also included as a control variable. Age is also included as a variable to represent the potential of a player. Being transferred in winter or summer also affects playing time, especially so in the season immediate after transfer, so this variable is included as an indicator.

Table 1 presents the summary all the variables. That standard deviations are much larger than the mean for some variables mean these variables are skewed heavily to the left, so there are many observations with small transfer fee and low playing time. The reason is due to that many of the observations include cases where a soccer club transfers players from its own youth team, which skews the data. Hence, there is perform appropriate robustness using observations with strictly positive transfer fee to alleviate such issues.

3.2 Description of Empirical Models for Sunk Cost Fallacy

I construct an unbalanced panel data of 6 seasons in total, including the season right after transfer up to the 5th season after a transfer. As many players transfer again, retire, or never gets playing time after a particular season, the number of observation decreases for each season after transfer. In the season right after transfer there are over 8000 forwards to observe, but the number decreases to 4000 for the season after, and eventually to 200 for the

Table 1: Summary Statistics of Key Variables

	Mean	Std. Dev.	Min	Max	Observations	Unique
Transfer Fee with Inflation	$2.514*10^6$	$7.896*10^6$	0	$2.289*10^8$	10087	
Total Minutes Played	1091.720	920.842	1	3986	24029	
Goal Efficiency	$2.984*10^{-3}$	$5.489*10^{-3}$	0	0.333	24029	
Assist Efficiency	$1.317*10^{-3}$	$4.968*10^{-3}$	0	0.5	24029	
Age	24.706	4.233	15	41	17929	
Points Earned per Appearance	1.265	0.609	0	3	24029	
Position					17929	4
Time of Transfer					10090	2
Team ID					32088	595

Note: Only forwards are included in the summary since they are the players to use for regression. Only players who have played are in the summary since goal and assist efficiency cannot be constructed with zero-minute playing time. "Unique" means the number of unique observations and is only applicable to discrete variables for fixed effects. The number of observations are not consistent in all variables because different variables may be missing for some players-season observations. Only players-season variables with all variables available are included for regression analysis.

5th season after transfer (Table 2). The rate of decrease in sample size is consistent with findings of Hackinger (2019).

For each of the t^{th} season after a transfer (the season right after transfer is counted as the 0th), I first construct an OLS regression with the following equation for player i:

$$minute_{(i,t)} = \alpha_{(i,t)} + \beta^{tr}_{(i,t)} ln(transfer_{(i,0)} + 1) + \beta^{sgo}_{(i,t)} \sqrt{goal_{(i,t)}} + \beta^{go}_{(i,t)} goal_{(i,t)} + \beta^{as}_{(i,t)} assist_{(i,t)} + \beta^{ag}_{(i,t)} age_{(i,t)} + \beta^{pt}_{(i,t)} point_{(i,t)} + \sum_{j}^{J} \beta_{(t,j)} team_{(i,t,j)} + \sum_{k}^{K} \beta_{(t,k)} position_{(i,t,k)} + \beta^{wi}_{(i,t)} winter_{(i,t)} + \epsilon_{(i,t)}. \quad (1)$$

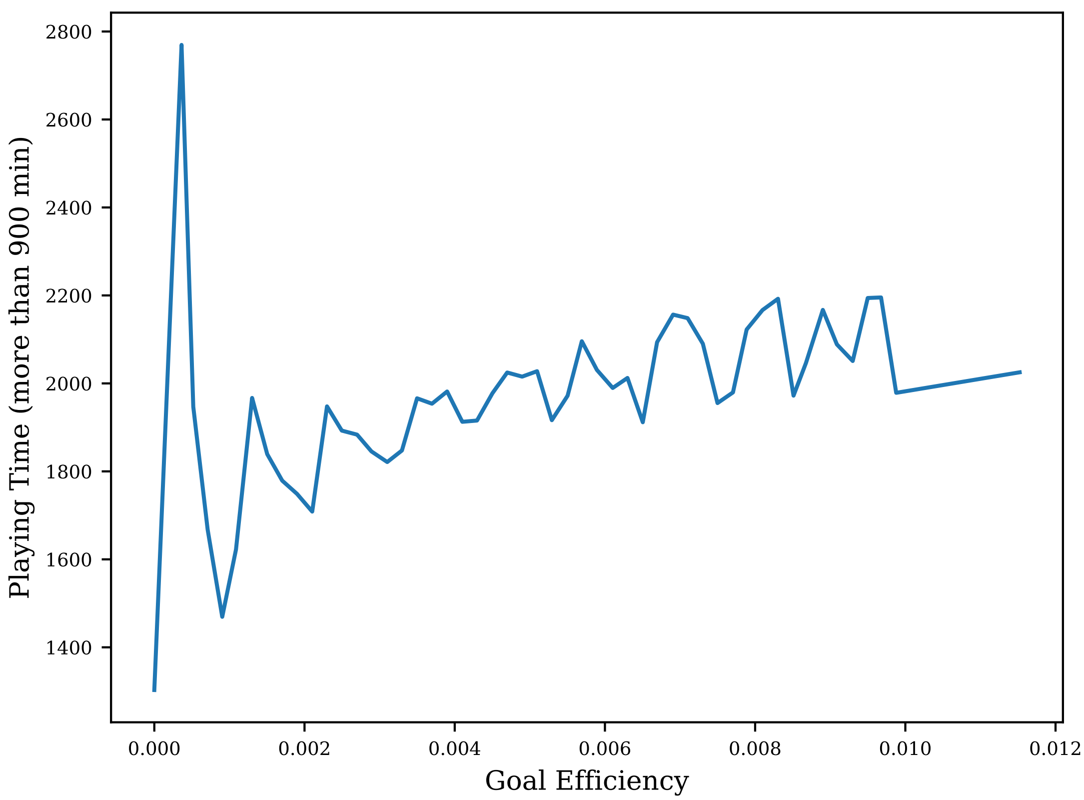
$minute_{(i,t)}$ is the playing time of player i in the t^{th} season ($t = 0, ..., 5$) after the season the transfer occurs. It is not possible that a transfer occurs after a player plays on a team, so transfer always occurs in season 0, which is denoted with $transfer_{(i,0)}$; I log this variable and, since some transfers have no cost, add 1 to $transfer_{(i,0)}$ to avoid infinity issues. $goal_{(i,t)}$ and $assist_{(i,t)}$ are performance control of player i in the t^{th} season after transfer using goal and assist efficiency.An additional issue with the most crucial control, goal efficiency, is that it is unlikely to have a linear relationship with playing time. The reason is that the bound of goal efficiency is very high compared to that of playing time. Moreover, it becomes more difficult to maintain a high goal efficiency once a player plays more. Therefore, I graphically observe the relationship between goal efficiency and playing time of all players (Figure 1). Players' goal efficiency and players' playing time of each season observed are broken into groups of goal efficiency within 0.0002 (e.g. goal efficiency from 0 to 0.0002 is group 1, 0.0002 to 0.0004 is group 2 etc.) up to goal efficiency of 0.01 and above. The data point is generated using average of goal efficiency for each group on x-axis and average of playing time on y. To avoid anomalies, only players who played more than 900 minutes in a given season (about 10 full games) are included.

From Figure 1, the rate at which playing time increases with goal efficiency plateaus (i.e. the marginal rate with which playing time increases with goal efficiency decreases).

Indeed, in section 4, adding square root of goal efficiency as a control results in a much higher R-squared compared with only using goal efficiency (Table 3; Table 4). Hence, I also add square root of goal efficiency as a control.

$age_{(i,t)}$ is the age of the player i in season t. $point_{(i,t)}$ is the number of points a player earns for the team every time he plays. $team_{(i,t,j)}$ and $position_{(i,t,k)}$ are dummy variables where $team_{(i,t,j)}$ is 1 if player i is in team j in the t^{th} season after transfer and 0 otherwise, while $position_{(i,t,j)}$ is 1 if player i is registered in transfermarkt as position k in the t^{th} season after transfer and 0 otherwise. $winter_{(i,t)}$ is a dummy variable that is 1 if player i is transferred in winter and 0 otherwise. Finally, $\epsilon_{(i,t)}$ is the error term for player i in season t. $a_{(i,t)}$ is the intercept for the regression with respect to the t^{th} season after transfer and the set of coefficient β is the affect of the particular variable to which it is attached; the superscripts are for notation purpose and are not exponents.

Figure 1: Observing the Relationship between Playing Time and Goal Efficiency



The assumption is that the team should maximize the team's performance, so they should play players according to their contribution on the field. In particular, they should give players time according to the performance indices, which are goal and assist efficiency. Since the main purpose for a player on the field is to make as many goals as possible, goal and assist efficiency are most representative of the players' performance. Therefore, the coefficient on transfer fee is exogenous to player performance, so it is considered as correlation between sunk cost and playing time controlled on performance. As a result, the null hypothesis corresponds with $\beta^{tr}_{(i,t)} \leq 0$, while the alternate hypothesis corresponds with $\beta^{tr}_{(i,t)} > 0$.

Since there is significant loss of number of observations as time progresses, there are likely noise generated by players whom the teams do not commit to in earlier seasons; this subset of observations likely cause bias. As a result, there is additional OLS regression where only players who played in a team for at least two seasons are kept. As discussed with the variable summary, there is an overwhelming amount of free transfers, which causes biases as the data is heavily skewed towards the left. Meanwhile, there is also some large discontinuity between free transfers and transfers paid with a fee, so there is also more robustness checks where only forwards with a strictly positive transfer fee are kept.

In addition, this paper checks coefficient stability by applying method developed by Altonji et al. (2000) and Oster (2017). Specifically, Oster (2017) expands on Altonji et al. (2000) to derive the following equation to generate coefficient that eliminates unobserve variable bias:

$$\beta^* = \hat{\beta} - \delta(\beta^o - \hat{\beta})\frac{R_{max} - \hat{R}}{\hat{R} - R^o} \quad (2)$$

In this case, β refers to $\beta^{tr}_{i,t}$, where $t = 0, ..., 5$, as transfer fee is the independent variable of interest (Model 1). In the equation, $\hat{\beta}$ is the coefficient on the log transfer fee with all observed controls. β^o is the coefficient on the log transfer fee without any control. $\hat{R}$ is R-squared from the regression with control and R^o is that from the univariate regression. R_{max} is the hypothetical R-squared if all the controls (observed and unobserved) are included in the regression, and δ measures how strong the observed variable correlates with log transfer fee relative to that of unobserved variables. In particular, using W_1 to represent the set of observed variables and W_2 to represent the set of unobserved variables, δ must satisfy the following relationship: $\delta\frac{cov(W_1, log(transfer\ fee+1))}{var(W_1)} = \frac{cov(W_2, log(transfer\ fee+1))}{var(W_2)}$. Oster (2017) proves that β^* converges to the 'true' β without unobserved variable bias.

In order to calculate β^*, all variables in Model 2 are known except R_{max} and δ. The maximum possible value R_{max} may take is 1, so $R_{max} \leq 1$. In addition, although the unobserved variables are unknown, the controls already obtained are comprehensive. In particular, there are measures for how well a forward scores (goal efficiency), how well he helps his team mates to score (assist efficiency), and how well he contributes to team wins (points earned per appearance). In addition, I control for environment and potential of players using team fixed effects and age. Although these variables do not cover all controls, they are comprehensive in that they include all essential performance and team factors. The rest of the variables that cover characteristics are unlikely to correlate more with log transfer fee than the observed control do because these variables are partially unobserved by teams purchasing players, so a) they do not correlate as strong with transfer fee; and b) the most important qualities of players are performance and contribution to wins, which is controlled for by observed variables. Hence, I assume $\delta \leq 1$, so $\beta^* \leq \hat{\beta} - (\beta^o - \hat{\beta})\frac{1-\hat{R}}{\hat{R}-R^o}$. I use the right handside of the inequality, which is Model 2 with $\delta = R_{max} = 1$, to estimate $\beta^\star$ for a lower bound of $\beta^{tr}_{i,t}$. Since players are observed for six seasons after transfer, there are six $\beta^{tr}_{i,t}$, and the above equation may be used to estimate each one of them.

Lastly, the paper also attempts to test for causation by using information from the team selling players. The paper uses selling team balance as instrument for transfer fee. The teams that have the best players do not necessarily have a positive or negative balance. It does not even necessarily have a large balance (in terms of absolute value) as a team may choose to equalize its revenue and cost, thus having balance close to zero. Hence, players quality, which may link to how good a club is, tends to be unrelated to the balance of the selling team. However, a positive balance tends to indicate that a team is good at negotiating players or need more money for financial reasons. Hence, higher balance correlates with transfer fee positively as teams with higher balance tends to have the skills and incentives to bargain for higher transfer fee. Indeed, Table A.1 and Table A.2 show that balance of the selling team is independent of the performance players after transfer.

For the instrumental variable regression, the same controls as the OLS method are used, so the first stage regression is:

$$\begin{aligned}\widehat{ln(transfer_{(i,0)}+1)} = \alpha_{(i,t)} + \beta^{ba}_{(i,0)} ln(balance+1) + \beta^{sgo}_{(i,t)}\sqrt{goal_{(i,t)}} + \beta^{go}_{(i,t)} goal_{(i,t)} + \\ \beta^{as}_{(i,t)} assist_{(i,t)} + \beta^{ag}_{(i,t)} age_{(i,t)} + \beta^{pt}_{(i,t)} point_{(i,t)} + \\ \sum_{j}^{J} \beta_{(t,j)} team_{(i,t,j)} + \sum_{k}^{K} \beta_{(t,k)} position_{(i,t,k)} + \beta^{wi}_{(i,t)} winter_{(i,t)} + \epsilon_{(i,t)} \quad (3)\end{aligned}$$

where logarithm of selling team balance of the same year that the player is transferred is the instrument. The above first stage regression is run for six times since I use instrumental variable technique for playing time for six years after transfer. The second stage regression is then:

$$\begin{aligned}minute_{(i,t)} = \alpha_{(i,t)} + \beta^{tr}_{(i,0)} \widehat{ln(transfer_{(i,0)}+1)} + \beta^{sgo}_{(i,t)}\sqrt{goal_{(i,t)}} + \beta^{go}_{(i,t)} goal_{(i,t)} + \\ \beta^{as}_{(i,t)} assist_{(i,t)} + \beta^{ag}_{(i,t)} age_{(i,t)} + \beta^{pt}_{(i,t)} point_{(i,t)} + \\ \sum_{j}^{J} \beta_{(t,j)} team_{(i,t,j)} + \sum_{k}^{K} \beta_{(t,k)} position_{(i,t,k)} + \beta^{wi}_{(i,t)} winter_{(i,t)} + \epsilon_{(i,t)} \quad (4)\end{aligned}$$

where $\beta^{tr}_{(i,0)}$ being strictly positive means teams commit to sunk cost fallacy. For the entire two stage least square, the observations restrict to players with strictly positive transfer fee since free transfer or transfer from youth league has no correlation with the status of the selling team (since the selling team does not have a say in transfer fee in this case).

3.3 Description of Empirical Models for Recency Bias

For recency bias, the effect cannot be measured through one single regression. Instead, I evaluate whether teams weigh recent performance prior to transfer too heavily by multiple regressions using performance prior to transfer as the independent variables and transfer fee and performance after transfer as the dependent variables.

I first evaluate how much teams weigh performance prior to transfer to determine transfer fee:

$$
\begin{aligned}
ln(transfer_{(i,0)}+1) = \alpha_{(i,0)} + \sum_{s=1}^{3}\beta^{(go,tr,0)}_{(i,-s,0)} goal_{(i,-s)} + \sum_{s=1}^{3}\beta^{(pt,tr,0)}_{(i,-s,0)} point_{(i,-s)} \\
+ \sum_{s=1}^{3}\sum_{c}^{C}\beta^{(tr,comp,0)}_{(i,-s)} competition_{(i,-s,c)} + \beta^{(ag,tr,0)}_{(i,0)} age_{(i,0)} + \sum_{j}^{J}\beta^{(tr,0)}_{(t,j)} team_{(i,0,j)} \\
+ \sum_{k}^{K}\beta^{(tr,0)}_{(0,k)} position_{(i,0,k)} + \beta^{(wi,tr,0)}_{(i,0)} winter_{(i,0)} + \epsilon_{(i,0)}. \quad (5)
\end{aligned}
$$

The variables transfer, goal, point, age, team, position, and winter appear in Model 1 and have the same meaning. However, the difference is in the season -s. The performance variables, goal and point, are goal efficiency and points earned per appearance of players prior to their first season after transfer. Hence, $-s$ means s season prior to transfer, so $goal_{(i,-s)}$ is the goal efficiency of player i in season $-s$, which is s season prior to transfer of player i. Goal efficiency is used instead of the square root of goal efficiency because, unlike playing time, $ln(transfer\ fee + 1)$ correlates with goal efficiency linearly because a) tranfer fee is linearized using logarithm; and b) transfer fee, unlike playing time, is theoretically unbounded. Hence, goal efficiency is more suitable than its square root as the independent variable in this case.

Another issue with goal efficiency is that players may have obtained multiple versions of this variable in the same season when competing in multiple competitions. In this case, only the player's performance in a league (but not necessarily limited to the top five league), and in the case that a player may have competed for multiple teams, I include the observation for which players have played more as evidenced by higher total minutes played to generate the variables goal and point. Moreover, performance indicating variables in different competitions have very different indication as to how players may perform after transfer. It is much more difficult to have high goal efficiency in a competitive league like La Liga or the Premier league than to have high scoring efficiency in a less competitive league. Hence, a new variable $competition_{(i,-s,c)}$ is used to control for competition c, in which player i was playing s season prior to his transfer. It is ideal to control even for the specific team the player played for. However, there are so many teams that a player may be from, and there are too many cases where there is only one direct transfer between a selling team and a team purchasing a player. Hence, I instead control for competition.

An almost identical regression is used to evaluate for goal efficiency, or player performance, for the first three seasons after player transfers:

$$goal_{i,t} = \alpha_{(i,t)} + \sum_{s=1}^{3} \beta_{(i,-s)}^{(go,go,t)} goal_{(i,-s)} + \sum_{s=1}^{3} \beta_{(i,-s)}^{(pt,go,t)} point_{(i,-s)} + \sum_{s=1}^{3}\sum_{c}^{C} \beta_{(i,-s)}^{(comp,go,t)} competition_{(i,-s,c)} + \beta_{(i,t)}^{(ag,go,t)} age_{(i,t)} + \sum_{j}^{J} \beta_{(t,j)}^{(go,t)} team_{(i,t,j)} + \sum_{k}^{K} \beta_{(t,k)}^{(go,t)} position_{(i,t,k)} + \beta_{(i,t)}^{(wi,go,t)} winter_{(i,t)} + \epsilon_{(i,t)}. \quad (6)$$

It is important to distinguish between the dependent variable $goal_{(i,t)}$ and the control variable $goal_{(i,-s)}$ since the former measures goal efficiency after transfer in the new team while the latter measures goal efficiency before transfer.

Free transfers or signing from youth teams, in this case, are not good indication of how much team are willing to pay as teams do not actively determine the price to pay for players. Hence, only players with a strictly positive transfer fee and who have played for at least three seasons prior to transfer are included in Model 5 and Model 6.

Since performance for players for the first three seasons after transfer are evaluated, Model 6 is evaluated three times for $t = 0, 1, 2$. The only difference between Model 5 and Model 6 is the dependent variable. Therefore, if teams do not over-weigh players' recent performance, then the relative relationship between the coefficients of $goal_{i,-s}$ in Model 5 and Model 5 for $s = 1, 2, 3$ should be similar. In this case, teams accurately predict how each year in the past may influence future performance and use the accurate prediction to determine transfer fee. On the other hand, if the extent that $\beta_{(i,-1)}^{(go,tr,0)}$ is greater than $\beta_{(i,-2)}^{(go,tr,0)}$ and $\beta_{(i,-3)}^{(go,tr,0)}$ is relatively higher than the extent that $\beta_{(i,-1)}^{(go,go,t)}$ is greater than $\beta_{(i,-2)}^{(go,go,t)}$ and $\beta_{(i,-3)}^{(go,go,t)}$, then teams are placing too much emphasis on recent performance as the relative impact of recent performance on future performance is smaller.

To quantify the comparison between relative impact of past performance on transfer fee and that on future performance, I compute the ratio between coefficients as most researchers studying recency bias in sports do (Fox, 2015; Healy, 2008; Hochberg, 2011). More specifically, I compare the ratios $\frac{\beta_{(i,-2)}^{(go,tr,0)}}{\beta_{(i,-1)}^{(go,tr,0)}}$ and $\frac{\beta_{(i,-3)}^{(go,tr,0)}}{\beta_{(i,-1)}^{(go,tr,0)}}$ with the ratios $\frac{\beta_{(i,-2)}^{(go,go,t)}}{\beta_{(i,-1)}^{(go,go,t)}}$ and $\frac{\beta_{(i,-3)}^{(go,go,t)}}{\beta_{(i,-1)}^{(go,go,t)}}$ respectively for $t = 0, 1, 2$. The standard errors of each ratio is generated through bootstrapping of 1000 iterations of sampling with replacement for Model 5 and 6. If the coefficients are statistically different with one-tail T-test for the ratio of coefficients in Model 6 being larger than that in Model 5, then the teams commit to recency bias.

In the analysis for recency bias, only goal efficiency is used as the performance indication after transfer because the other performance variable, assist efficiency, is statistically insignificant in determining playing time (Table 2), so teams do not weigh other performance as heavily as goal efficiency. Moreover, the other key performance variable, points earned per appearance, is mostly statistically insignificant for even the immediate season prior to transfer (on goal efficiency after transfer and on points earned per appearance after transfer), which causes the issue that the ratio of coefficients on points earned per performance is not defined. Hence, only use goal efficiency is used for the analysis.

4 Result and Interpretations

4.1 Sunk Cost Fallacy Ordinary Least Square

I first present the ordinary least square regression (OLS) by directly regressing playing time on the variables from Model 1. The fixed effects are taken but not presented in the table for length concerns.

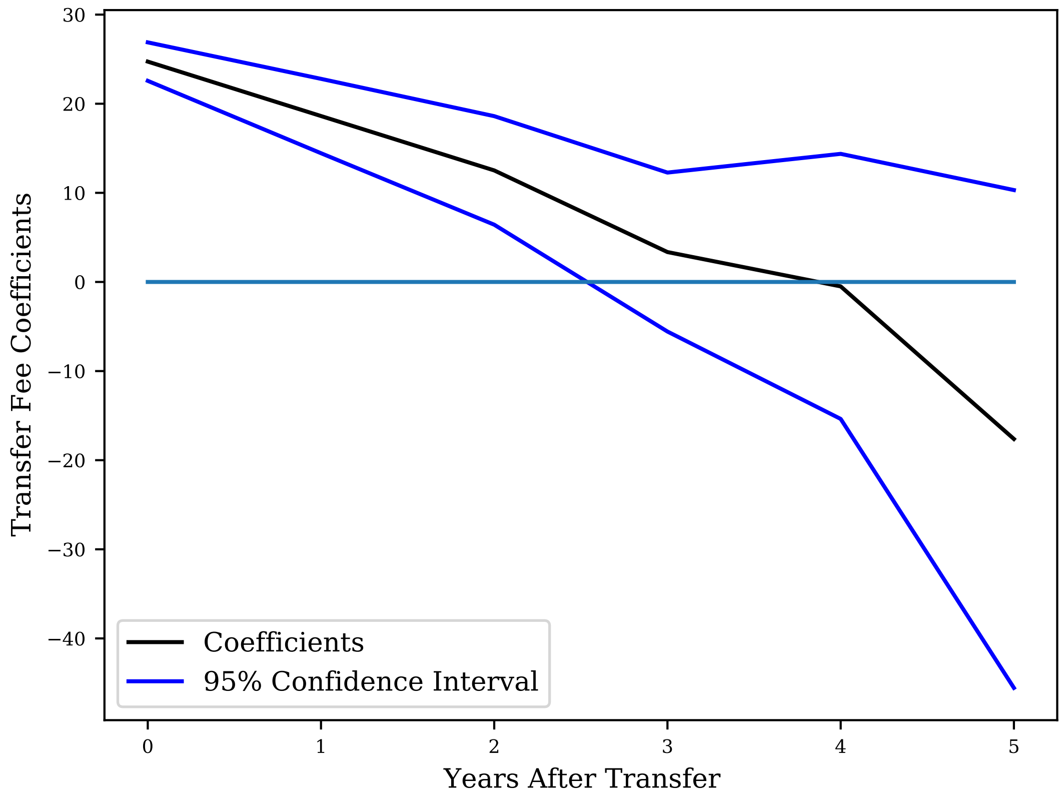
Table 2: Regression for Playing Time with Square Root Goal Efficiency

	Playing Time Year of Transfer	Playing Time 1 Year from Transfer	Playing Time 2 Year from Transfer	Playing Time 3 Year from Transfer	Playing Time 4 Year from Transfer	Playing Time 5 Year from Transfer
	(1)	(2)	(3)	(4)	(5)	(6)
log(Transfer Fee + 1)	24.728***	18.620***	12.516***	3.355	-0.497	-17.620
	(1.079)	(2.086)	(3.045)	(4.460)	(7.436)	(13.966)
$\sqrt{\text{Goal Efficiency}}$	16486.513***	22174.744***	27557.546***	28652.391***	40110.952***	39348.193***
	(342.481)	(743.444)	(1187.494)	(1886.457)	(4929.059)	(8405.384)
Goal Efficiency	-65568.878***	-99303.161***	-133175.433***	-151777.196***	-252105.490***	-227640.772***
	(2376.047)	(5208.707)	(8208.347)	(13842.987)	(54715.857)	(78573.031)
Assist Efficiency	479.770	4104.485	-3358.984	25371.997***	-16848.135	18614.922
	(1428.822)	(4971.125)	(6398.783)	(7741.216)	(22286.522)	(50158.105)
Age	22.786***	12.319***	-1.786	-2.679	-34.885**	-19.875
	(1.817)	(3.619)	(5.620)	(8.714)	(14.390)	(26.472)
Points Earned per Appearance	114.132***	159.294***	105.319**	163.544**	354.157***	-11.891
	(13.560)	(30.442)	(45.380)	(68.890)	(114.434)	(177.024)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8899.0	3357.0	1685.0	859.0	402.0	209.0
R2	0.387	0.36	0.417	0.561	0.752	0.934
Adjusted R2	0.342	0.232	0.199	0.231	0.242	0.468
Residual Std. Error	676.564(df = 8284.0)	808.717(df = 2795.0)	841.001(df = 1226.0)	824.915(df = 490.0)	802.781(df = 131.0)	688.379(df = 26.0)
F Statistic	8.517***	2.804***	1.911***	1.699***	1.474***	2.006**)
	(df = 614.0; 8284.0)	(df = 561.0; 2795.0)	(df = 458.0; 1226.0)	(df = 368.0; 490.0)	(df = 270.0; 131.0)	(df = 182.0; 26.0)

*p<0.1; **p<0.05; ***p<0.01

Recalling from Model 1, the result shows that the term $\beta^{tr}_{(i,t)}$, which is the coefficient for log of transfer fee, is positive at five percent statistically significant level for t=0,1,2, meaning that, controlled on performance level, higher transfer fee correlates with more playing time for the first two years after a player is transferred. The interpretation for coefficient $b^{tr}_{(i,t)} = x$ is that doubling the transfer results in x more minutes of playing time in the t^{th} season after transfer. The coefficient is 25 for year 0, 19 for year 1, 13 for year 2, and 3 (which is not statistically significant) for year 3 (Table 2), showing decrease in the effect of transfer fee on playing time as time progresses (Figure 2). The decrease in the economic magnitude of the coefficient on the independent variable is consistent with all the literature that found presence of sunk cost fallacy of sports teams (Hinton & Sun, 2020; Keefer, 2015, 2017; Staw & Hoang, 1995). The result shows that the higher transfer fee does correlate with more playing time for the first three years after transfer, but the null hypothesis is true for the fourth to sixth year after transfer. Such a result is already different from the findings

Figure 2: Plot of Coefficients on log(Transfer Fee + 1)



of Hackinger (2019), who did not find significant effect transfer fee has on playing time with OLS results.

However, it is possible that the coefficients for log of transfer fee are biased. The first reason is that there is high possibility of omitted variable bias. While goal and assist efficiency and age are used as controls, a player may have other intrinsic qualities—such as leadership, training quality, popularity among fans etc.—that are unrelated to performance but causes a team to play the player more, which refers to the possibility that transfer fee may positively correlate with the unobserved variables. Therefore, even though there is a positive correlation between transfer fee and playing time controlled on performance, it is likely that the coefficient is biased upward. Also, the data obtained is heavily skewed. There are many players with free transfer fee, and these are more likely players who have relatively lower abilities and thus lower playing time. It is possible that high concentration of these observations have skewed the data and thus increased the coefficient. To address these concerns, I demonstrate coefficient stability after taking omitted variable bias into account in Section 4.2 and address the issue of skewed data in Section 4.3.

4.2 Sunk Cost Fallay Coefficient Stability

I use the method introduced by Altonji et al. (2000) and Oster (2017) to examine whether the coefficients obtained from Table 1 are still positive without unobserved variable bias. In other words, there is need to test whether sunk cost fallacy persists when all control variables–which are observed or unobserved–relevant to playing time are included. Since the coefficients on $log(transfer + 1)$ stands to be robust for all robustness checks for year of transfer and first year after transfer (Table 6; Table 7), these years are investigated more detailedly by presenting all regressions when observed control variables are gradually included.

Table 3: All Regressions for Playing Time in Year of Transfer

	Dependent variable: Playing Time Year of Transfer for All Columns					
	Univariate	Fixed Effects	Goal Efficiency	Assist Efficiency	All Other Controls	Sqaure Root of Goal Efficiency
	(1)	(2)	(3)	(4)	(5)	(6)
log(Transfer Fee + 1)	31.573***	37.777***	36.584***	36.551***	34.514***	24.728***
	(1.178)	(1.242)	(1.232)	(1.232)	(1.193)	(1.079)
$\sqrt{\text{Goal Efficiency}}$						16486.513***
						(342.481)
Goal Efficiency			24919.727***	24911.544***	20169.454***	-65568.878***
			(1814.976)	(1814.840)	(1770.120)	(2376.047)
Assist Efficiency				2551.032	-378.832	479.770
				(1655.478)	(1608.061)	(1428.822)
Age					34.986***	22.786***
					(2.025)	(1.817)
Points Earned per Appearance					248.084***	114.132***
					(14.937)	(13.560)
Position FE	No	Yes	Yes	Yes	Yes	Yes
Team FE	No	Yes	Yes	Yes	Yes	Yes
Observations	8,972	8,899	8,899	8,899	8,899	8,899
R^2	0.074	0.177	0.194	0.194	0.249	0.407
Adjusted R^2	0.074	0.157	0.175	0.175	0.230	0.392
Residual Std. Error	801.284(df = 8970)	765.661(df = 8687)	757.529(df = 8686)	757.469(df = 8685)	731.460(df = 8681)	649.879(df = 8680)
F Statistic	718.174***	8.835***	9.873***	9.839***	13.272***	27.366***
	(df = 1.0; 8970.0)	(df = 211.0; 8687.0)	(df = 212.0; 8686.0)	(df = 213.0; 8685.0)	(df = 217.0; 8681.0)	(df = 218.0; 8680.0)

*p<0.1; **p<0.05; ***p<0.01

Table 3 and Table 4 demonstrate that adding in the controls gradually decreases the effect of log(Transfer Fee + 1) on playing time while the R-squared increases. With all observed control variables, the coefficient for both year of transfer and one year after transfer remain statistically significant. However, the question is, if more controls (the unobserved variables) are included, will the coefficient decreases to the extent that the coefficient will become non-positive? In other words, if more controls are added, will sunk cost fallacy disappear by this trend of decrease in coefficient?

I attempt to answer this question through two ways. The first way is to graphically interpret the trend. I graph the coefficient of log(transfer+1) in the regressions in Table 3 and Table 4 on the y-axis and the R-squared of these regressions on the x-axis to analyze the speed through which the coefficient decreases while R-squared increases. R-squared is bounded by 1, so if the trend shows that the coefficient still remains positive when R-squared is 1, it suggests that the coefficient is likely positive when all the controls are included.

By drawing linear best fit through the points in Figure 3 and Figure 4, in both cases,

Table 4: All Regressions for Playing Time one Year after Transfer

	Dependent variable: Playing Time 1 Year after Transfer for All Columns					
	Univariate	Fixed Effects	Goal Efficiency	Assist Efficiency	All Other Controls	Sqaure Root of Goal Efficiency
	(1)	(2)	(3)	(4)	(5)	(6)
log(Transfer Fee + 1)	25.993***	30.417***	29.794***	29.656***	28.153***	18.620***
	(2.044)	(2.389)	(2.371)	(2.371)	(2.335)	(2.086)
$\sqrt{\text{Goal Efficiency}}$						22174.744***
						(743.444)
Goal Efficiency			27687.547***	27656.714***	18863.173***	-99303.161***
			(3801.877)	(3799.919)	(3829.437)	(5208.707)
Assist Efficiency				11822.953**	5406.607	4104.485
				(5703.269)	(5629.197)	(4971.125)
Age					9.739**	12.319***
					(4.097)	(3.619)
Points Earned per Appearance					361.102***	159.294***
					(33.611)	(30.442)
Position FE	No	Yes	Yes	Yes	Yes	Yes
Team FE	No	Yes	Yes	Yes	Yes	Yes
Observations	3,377	3,357	3,357	3,357	3,357	3,357
R^2	0.046	0.106	0.121	0.122	0.156	0.342
Adjusted R^2	0.045	0.049	0.064	0.065	0.100	0.298
Residual Std. Error	901.910(df = 3375)	899.810(df = 3153)	892.476(df = 3152)	892.009(df = 3151)	875.321(df = 3147)	772.963(df = 3146)
F Statistic	161.675***	1.850***	2.131***	2.144***	2.783***	7.789***
	(df = 1.0; 3375.0)	(df = 203.0; 3153.0)	(df = 204.0; 3152.0)	(df = 205.0; 3151.0)	(df = 209.0; 3147.0)	(df = 210.0; 3146.0)

*p<0.1; **p<0.05; ***p<0.01

the trend shows that the coefficient on log of transfer fee is above zero when R-squared is 1, suggesting that sunk cost fallacy likely remains even more unobserved variables are included as control. However, the graphical method is not mathematically robust since switching the order that the controls are added can change the coordinates in the graph. If I were to add in the controls in different orders in Table 3 and Table 4, the coordinates will change in Figure 3 and Figure 4, which changes the best fit.

For the same reason, fitting with polynomials of degree higher than one is meaningless because changing the order of the controls can have drastic different effects. By visually examining the coordinates in Figure 3 and Figure 4, coefficient seems to accelerate to hit the zero line before R-squared reaches 1. However, if square-root of goal efficiency were included as control first (which decreases the coefficient significantly) and the fixed effects are included last (which increases the coefficient), the coefficients would seem to be accelerating upward.

Therefore, even if coefficient on log of transfer fee is above zero graphically, there is still need to confirm coefficient stability through calculating the coefficient when unobserved variables are included (Oster, 2017). Specifically, I use Equation 2 to calculate the 'true' coefficient of $\beta_{i,t}^{tr}$ for $t = 0, 1, ..., 5$ (Model 1) with the assumption that $R_{max} = 1$ and $\delta = 1$.

Table 5 shows that β^*, which is the true coefficient calculated using Equation 2, is positive for $\beta_{i,t}^{tr}$ for $t = 0, 1, 2$, which means that sunk cost fallacy are likely to persist even if all unobserved variables are included. Moreover, in section 3, I argued that assuming $R_{max} = 1$ and $\delta = 1$ means the coefficient β^* calculated is a lower bound for $\beta_{i,t}^{tr}$. Hence, the 'true' $\beta_{i,t}^{tr}$s are likely even higher than those in Table 5. Therefore, through both graphical and mathematical methods, I show that sunk cost fallacy likely persists even when unobserved variable bias is excluded.

Figure 3: Test of Coefficient Stability with Unobserved Variables Using Results from the Year of Transfer by Methods of Altonji et al. (2000) and Oster (2017)

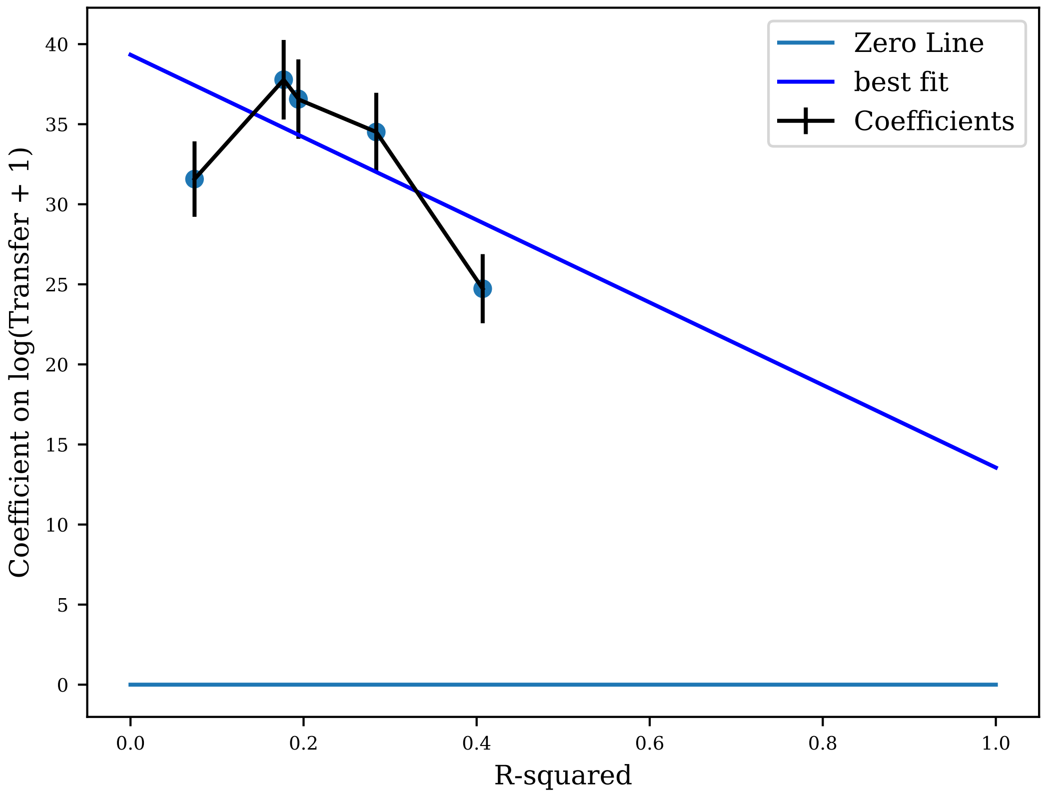


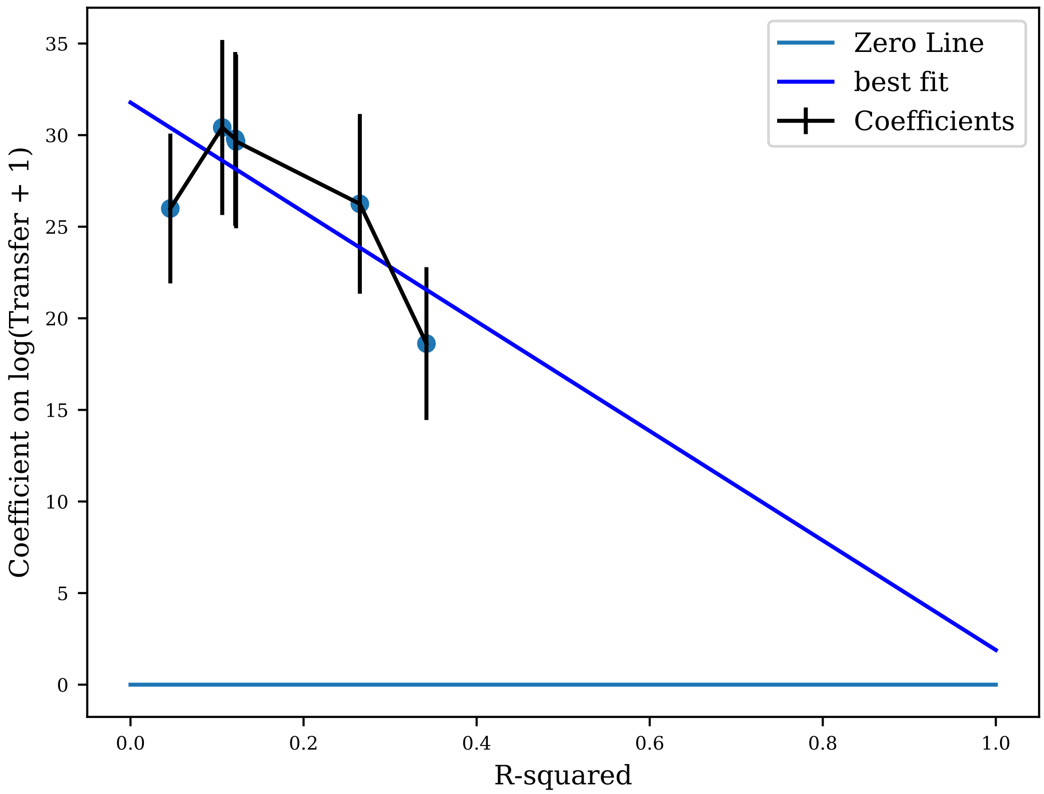
Table 5: 'True' Coefficient Calculation

	$\hat{R}$	R^o	$\hat{\beta}$	β^o	β^*
Year 0	0.407	0.074	24.728	31.573	12.539
Year 1	0.342	0.046	18.620	25.993	2.230
Year 2	0.400	0.017	12.516	16.254	6.660
Year 3	0.482	0.009	3.355	11.763	-5.853
Year 4	0.590	0.006	-0.497	9.488	-7.507
Year 5	0.685	0.001	-17.62	4.407	-27.764

4.3 Sunk Cost Fallacy Robustness Checks

To resolve the concern regarding free transfers skewing the data set, I attempt to use robustness tests with players with strictly positive transfer fee to alleviate the issue of skewed data

Figure 4: Test of Coefficient Stability with Unobserved Variables Using Results from One Year after Transfer by Methods of Altonji et al. (2000) and Oster (2017)



set and players who have played for a team for a certain amount of time. The regression with strictly positive transfer fee directly resolve the issue regarding high concentration of players with lower transfer fee. For the latter robustness check, I use the subset of players who played in season 2 after transfer for a team. In this case, the team must recognize and appreciate the abilities of these players in order to keep and play them for at least 3 seasons (season 0, 1, and 2), so there is no longer a concentration of player with low abilities and thus the data is no longer skewed to the left.

For the first robustness check with positive transfer fee, $b^{tr}_{(i,t)}$, the coefficient for transfer fee from Model 1, is still positive at five percent statistically significant level for i=0,1 with coefficient 104 for year 0, 81 for year 1, and 36 for year 2 (not statistically significant). The reason that the magnitude of coefficient is larger is likely due to that there are a lot of free transfers in the standard regressions, which reduces the variation in playing time and transfer fee, thus causing the coefficient to be smaller for regressions including all observations (Table 6).

For the second robustness check, the regression includes only players who have played in year 2 (thus staying for at least 2 years) and those who played in year 2 and had positive transfer fee (Table 7). For the observations that include zero transfer fee, the

Table 6: Robustness Check with Strictly Positive Transfer Fee

	Playing Time Year of Transfer	Playing Time 1 Year from Transfer	Playing Time 2 Year from Transfer	Playing Time 3 Year from Transfer	Playing Time 4 Year from Transfer	Playing Time 5 Year from Transfer
	(1)	(2)	(3)	(4)	(5)	(6)
log(Transfer Fee + 1)	103.862***	81.196***	35.741	-63.895	-53.357	-48.418
	(10.731)	(19.640)	(28.442)	(39.429)	(69.392)	(223.325)
$\sqrt{\text{Goal Efficiency}}$	18841.766***	36844.287***	27544.669***	25662.010***	28600.282***	23617.336
	(765.806)	(1681.173)	(1783.240)	(2608.510)	(8260.497)	(18478.723)
Goal Efficiency	-89569.479***	-261156.251***	-123999.355***	-127639.811***	-128235.009	-74678.115
	(6462.279)	(16349.408)	(11072.889)	(16261.812)	(84144.195)	(141391.307)
Assist Efficiency	-2341.753	2542.740	2519.956	18264.985**	-15191.055	38862.642
	(2509.097)	(8151.208)	(14171.750)	(8612.165)	(44549.864)	(81492.354)
Age	27.671***	0.446	-8.301	-14.558	-25.200	-11.634
	(3.715)	(6.122)	(9.336)	(13.694)	(25.467)	(59.916)
Points Earned per Appearance	132.501***	202.540***	42.460	169.173*	460.429**	-664.640
	(27.089)	(48.352)	(65.275)	(96.995)	(200.170)	(432.355)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,324	1,666	915	486	221	107
R^2	0.428	0.400	0.419	0.527	0.571	0.727
Adjusted R^2	0.390	0.324	0.292	0.331	0.193	-0.034
Residual Std. Error	662.511(df = 3117)	739.346(df = 1479)	782.729(df = 750)	753.689(df = 343)	795.026(df = 117)	904.378(df = 28)
F Statistic	11.322***	5.298***	3.297***	2.693***	1.510**	0.955
	(df = 206.0; 3117.0)	(df = 186.0; 1479.0)	(df = 164.0; 750.0)	(df = 142.0; 343.0)	(df = 103.0; 117.0)	(df = 78.0; 28.0)

*p<0.1; **p<0.05; ***p<0.01

coefficients are still positive at statistically significant level for year 0 to 2 with coefficient 18 for year 0, 13 for year 1, and 13 for year 2 (Table 7).

Table 7: Robustness Check with Players who Played in Year 2

	Playing Time Year of Transfer	Playing Time 1 Year from Transfer	Playing Time 2 Year from Transfer	Playing Time 3 Year from Transfer	Playing Time 4 Year from Transfer	Playing Time 5 Year from Transfer
	(1)	(2)	(3)	(4)	(5)	(6)
log(Transfer Fee + 1)	18.116***	13.395***	12.516***	-0.803	-5.127	-25.580
	(2.812)	(3.211)	(3.045)	(4.978)	(8.576)	(17.090)
$\sqrt{\text{Goal Efficiency}}$	27258.520***	33808.108***	27557.546***	26722.310***	38772.499***	39656.022***
	(1544.820)	(1840.229)	(1187.494)	(2128.932)	(5522.763)	(9536.157)
Goal Efficiency	-194903.440***	-225758.540***	-133175.433***	-140219.958***	-242821.630***	-234398.893***
	(14208.479)	(15299.089)	(8208.347)	(14619.294)	(60123.849)	(87647.265)
Assist Efficiency	11168.818	-2979.074	-3358.984	21417.912***	-18560.313	41293.244
	(9724.437)	(13162.853)	(6398.783)	(8023.893)	(25138.326)	(55397.208)
Age	59.778***	33.342***	-1.786	-9.950	-41.908**	-10.062
	(5.354)	(6.030)	(5.620)	(9.627)	(16.557)	(32.452)
Points Earned per Appearance	105.601***	145.323***	105.319**	173.768**	334.115**	-108.549
	(39.684)	(55.162)	(45.380)	(76.730)	(132.279)	(212.049)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,475	1,458	1,685	741	336	168
R^2	0.536	0.362	0.400	0.488	0.583	0.676
Adjusted R^2	0.471	0.264	0.317	0.333	0.312	0.205
Residual Std. Error	652.804(df = 1292)	751.341(df = 1264)	776.216(df = 1481)	756.880(df = 568)	763.131(df = 203)	832.368(df = 68)
F Statistic	8.211***	3.712***	4.855***	3.148***	2.152***	1.435*
	(df = 182.0; 1292.0)	(df = 193.0; 1264.0)	(df = 203.0; 1481.0)	(df = 172.0; 568.0)	(df = 132.0; 203.0)	(df = 99.0; 68.0)

*p<0.1; **p<0.05; ***p<0.01

The robustness again confirms the alternate hypothesis for the first one to three years after transfer and shows the null hypothesis to be true for the years after. The reason that the numbers of observations for year 0 and 1 are smaller than that in year 2 is due to that there are players who played in year 2 that did not play in year 0 and 1. In this case, the economic magnitude of the coefficient does not drop significantly, meaning that the skewedness of the data does not cause the coefficients to be significantly biased. Moreover, the trend that the economic magnitude decreases over time is also observed in all robustness checks.

In both robustness checks, reducing the issue of the data skewing toward the left does not hinder the positive correlation between transfer fee and playing time, which means that our alternate hypothesis is true for the first two or three seasons after a player transfers. Therefore, the issue regarding skewed data is not causing the coefficient on log transfer fee to be too high.

4.4 Sunk Cost Fallacy Instrumental Variable

After presenting suggestive evidence of sunk cost fallacy likely persisting with robustness checks and after eliminating unobserved variable bias, I test whether it is possible to argue for causation using instrumental variable regression. As mentioned in section 3, I use the log of selling team balance as the instrument since it likely correlates positively with log of transfer fee while is independent of playing time after transfer.

Table 8: First Stage Regression using Selling Team Balance as Instrument

	Transfer Fee in Year of Transfer	Transfer Fee from 1 Year Ago	Transfer Fee from 2 Years Ago	Transfer Fee from 3 Years Ago	Transfer Fee from 4 Years Ago	Transfer Fee from 5 Years Ago
	(1)	(2)	(3)	(4)	(5)	(6)
log(From Team Balance+1)	0.007***	0.002	0.003	-0.001	0.006	0.025**
	(0.001)	(0.002)	(0.003)	(0.004)	(0.007)	(0.009)
$\sqrt{\text{Goal Efficiency}}$	9.679***	9.382***	8.984***	7.007**	1.622	9.903
	(1.250)	(2.210)	(2.261)	(3.559)	(10.874)	(15.165)
Goal Efficiency	-26.154**	-40.745*	-46.830***	-31.594	45.773	-56.230
	(10.575)	(21.597)	(14.070)	(22.195)	(111.025)	(116.086)
Assist Efficiency	8.367**	7.763	-2.187	5.990	90.296	-63.117
	(4.168)	(10.790)	(18.147)	(11.569)	(58.555)	(72.229)
Age	-0.009	0.004	0.003	0.031	0.038	0.163***
	(0.006)	(0.008)	(0.012)	(0.019)	(0.034)	(0.046)
Points Earned per Appearance	-0.016	0.159**	0.145*	0.032	-0.304	-0.408
	(0.045)	(0.064)	(0.084)	(0.133)	(0.259)	(0.358)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,324	1,666	915	486	221	107
R^2	0.537	0.637	0.639	0.659	0.761	0.933

*p<0.1; **p<0.05; ***p<0.01

However, even though the correlation between log of balance of selling team and log of transfer fee is positive (with the exception of year 4) as predicted, the positive correlations are mostly statistically insignificant except for year 0 and year 6 (Table 8. The reason is likely due to decreasing number of observations as players transfer away from the buying team or retire. Nevertheless, balance of selling team is a weak instrument. Still, for the years that are robust to unobserved variable bias (year 0, 1, 2) in Table 5, the coefficients remain at least one standard error above zero with year 0 being statistically significant at 1%. Hence, the second stage regressions are still implemented.

Table 9: Second Stage Least Square Using Selling Team Balance as Instrument

	Playing Time Year of Transfer	Playing Time 1 Year from Transfer	Playing Time 2 Year from Transfer	Playing Time 3 Year from Transfer	Playing Time 4 Year from Transfer	Playing Time 5 Year from Transfer
	(1)	(2)	(3)	(4)	(5)	(6)
log(Transfer Fee+1)	-105.904	-243.032	-53.526	5,569.764	288.675	-883.998***
	(113.455)	(563.173)	(633.217)	(27,352.057)	(685.377)	(289.268)
$\sqrt{\text{Goal Efficiency}}$	20,954.908***	39,881.763***	28,339.755***	-13,897.223	28,267.797***	39,859.930***
	(1,325.513)	(5,554.752)	(5,902.422)	(192,587.502)	(6,538.365)	(12,079.236)
Goal Efficiency	-96,214.807***	-274,248.009***	-128,149.250***	51,068.699	-142,685.152*	-159,585.803*
	(7,149.192)	(28,497.570)	(31,334.384)	(869,089.838)	(79,160.833)	(89,553.996)
Assist Efficiency	-520.586	4,697.670	1,992.698	-17,795.865	-48,430.564	30,559.850
	(2,746.672)	(9,269.311)	(12,996.396)	(174,000.698)	(73,433.641)	(51,625.709)
Age	24.861***	1.130	-8.637	-192.005	-30.282	97.474*
	(4.090)	(6.457)	(8.489)	(866.871)	(27.863)	(51.367)
Points Earned per Appearance	127.856***	253.669**	53.284	-9.129	559.447**	-1,086.035***
	(27.893)	(102.442)	(107.210)	(1,072.944)	(268.932)	(305.713)
Position FE	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,324	1,666	915	486	221	107
R^2	0.357	0.289	0.411		0.475	0.570

Note: Missing R^2 in column (4) is due to negative R^2. *p<0.1; **p<0.05; ***p<0.01

Unfortunately, Table 9 shows that, with two stage least squares, the coefficients are statistically insignificant and are even in the negative direction. Moreover, the standard error are much larger not only than the coefficients in the OLS regression (Table 2) but also than those in Table 6 where used strictly positive transfer fees are used. Hence, using balance as instrument likely generated plenty of noise, which causes both the coefficients and the standard errors to be dissimilar to those in OLS regressions. Hence, in general, selling team variables did not generated results that support sunk cost fallacy due to fewer observations and limitations of the majority of the variables. Therefore, the paper does not find evidence that supports causation of higher sunk cost of transfer fee in increasing players' playing time.

4.5 Recency Bias Analysis

Finally, I present results generated regarding recency bias of soccer teams in payment of transfer fee using empirical methods of section 3.3. I first present the regressions for Model 5 and 6.

From the regressions in Table 10, the goal efficiency 3 years before transfer has significant positive correlation with goal efficiency in the first two years after transfer. However, this variable is a poor indicator for transfer fee. Hence, at a first glance, it seems that teams may indeed be overvaluing recent performance since performance in the distant past is quite important for performance after transfer but is insignificant in determining transfer fee.

However, for both transfer fee and performance after transfer, the importance of goal efficiency prior to transfer diminishes for years further in the past. Hence, teams correctly predict that performance in the more distant past matters less to performance after

Table 10: Preliminary Regressions for Recency Bias Analysis

	log(Transfer+1)	Goal Efficiency Year of Transfer	Goal Efficiency 1 Year after Transfer	Goal Efficiency 2 Years after Transfer
	(1)	(2)	(3)	(4)
Goal Efficiency 1 Year Before Transfer	81.335***	0.165***	0.239***	0.257*
	(10.026)	(0.024)	(0.043)	(0.150)
Goal Efficiency 2 Years Before Transfer	17.390***	0.030**	0.088**	-0.068
	(5.651)	(0.013)	(0.043)	(0.151)
Goal Efficiency 3 Years Before Transfer	2.641	0.023*	0.084**	0.043
	(5.460)	(0.013)	(0.040)	(0.139)
Points Earned per Appearance 1 Year Before Transfer	0.274***	$-0.125*10^{-3}$	$0.101*10^{-3}$	$-1.088*10^{-3}$
	(0.0592881)	$(0.139*10^{-3})$	$(0.235*10^{-3})$	$(0.816*10^{-3})$
Points Earned per Appearance 2 Year Before Transfer	0.118**	0.186^{-3}	-0.318^{-3}	0.026^{-3}
	(0.059)	$(0.139*10^{-3})$	$(0244*10^{-3})$	$(0.879*10^{-3})$
Points Earned per Appearance 3 Year Before Transfer	$2.689*10^{-3}$	$-0.130*10^{-3}$	$-0.009*10^{-3}$	$0.301*10^{-3}$
	$(51.475*10^{-3})$	$(0.120*10^{-3})$	$(0.213*10^{-3})$	$(0.759*10^{-3})$
Age	-0.075***	$-0.023*10^{-3}$	$-0.152*10^{-3***}$	$-0.127*10^{-3}$
	(0.009)	$(0.021*10^{-3})$	$(0.033*10^{-3})$	$(0.116*10^{-3})$
Position FE	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes
Competition FE	Yes	Yes	Yes	Yes
Observations	2,569	2,569	1,296	701
R^2	0.685	0.665	0.490	0.545
Adjusted R^2	0.589	0.563	0.213	0.082
Residual Std. Error	1.002(df = 1968)	0.002(df = 1968)	0.002(df = 839)	0.005(df = 347)
F Statistic	7.138***	6.516***	1.767***	1.178*
	(df = 600.0; 1968.0)	(df = 600.0; 1968.0)	(df = 456.0; 839.0)	(df = 353.0; 347.0)

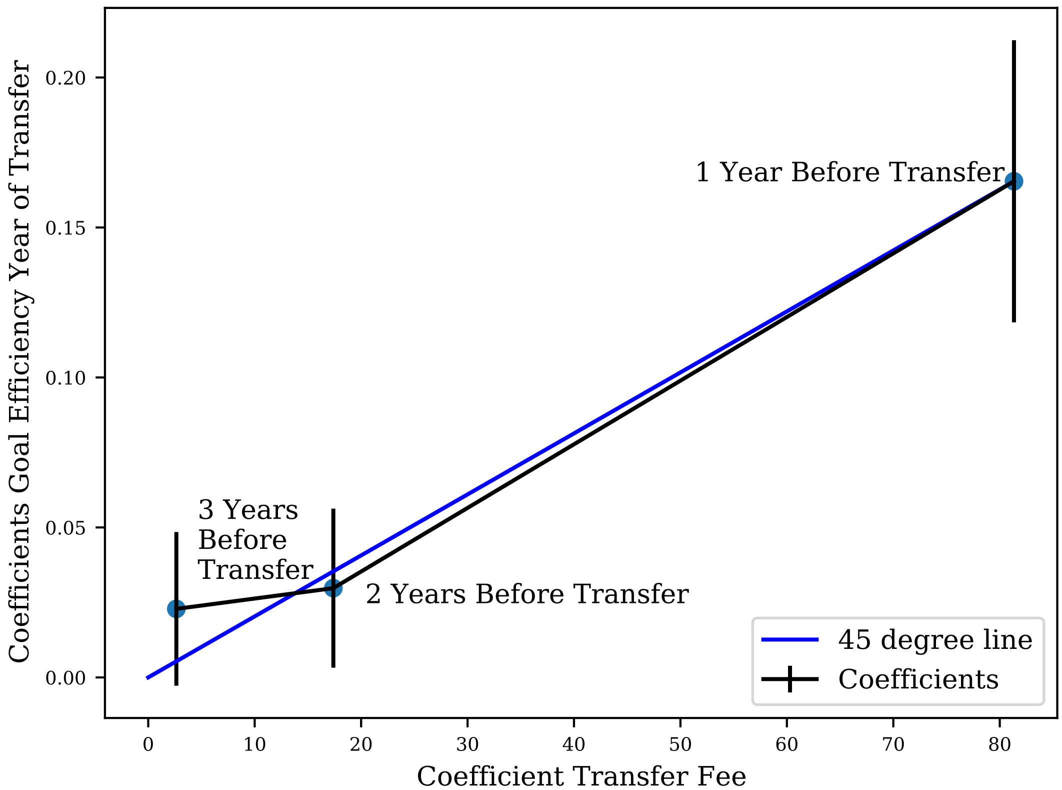
*p<0.1; **p<0.05; ***p<0.01

transfer. It is important to know, then, whether teams are over- or under-weighing the magnitude that the importance of performance in the more distant past diminishes. If teams correctly evaluate the amount of decrease of the importance of the past, then the coefficient on log(Transfer+1) should decrease proportionally at the same rate as the decrease for coefficients on performance after transfer.

In other words, referring to Model 5 and 6, the rate that $\beta^{(go,tr,0)}_{(i,-s)}$ decreases as s increases (from 1 to 3) should be the same rate that $\beta^{(go,go,t)}_{(i,-s)}$ decreases as s increases for $t = 0, 1, 2$. However, due to less observations, the coefficient in Table 10 column (4) are largely insignificant (which are for performance 2 years after transfer), so I mainly compare $\beta^{(go,tr,0)}_{(i,-s)}$ with $\beta^{(go,go,0)}_{(i,-s)}$ and $\beta^{(go,go,1)}_{(i,-s)}$, which are factors for performance during the year after transfer and one year after transfer.

To examine the comparison of the rate of decrease, I graph $\beta^{(go,go,0)}_{(i,-s)}$ vs. $\beta^{(go,tr,0)}_{(i,-s)}$ and $\beta^{(go,go,0)}_{(i,-s)}$ vs. $\beta^{(go,tr,0)}_{(i,-s)}$, which compares the magnitude of the weight of past performance on transfer fee with that on performance after transfer. If teams correctly predict the rate of decrease in the importance of performance in the more distant past, the coefficient should align with a linear interpolation with the origin and ($\beta^{(go,tr,0)}_{(i,-1)}$,$\beta^{(go,go,0)}_{(i,-1)}$) for Figure 5 and that with the origin and ($\beta^{(go,tr,0)}_{(i,-1)}$, $\beta^{(go,go,1)}_{(i,-1)}$,) for Figure 6, which are the 45 degree line drawn.

Figure 5: Test of Recency Bias Using Relative Weights of Coefficients of Past Performance from the Year of Transfer

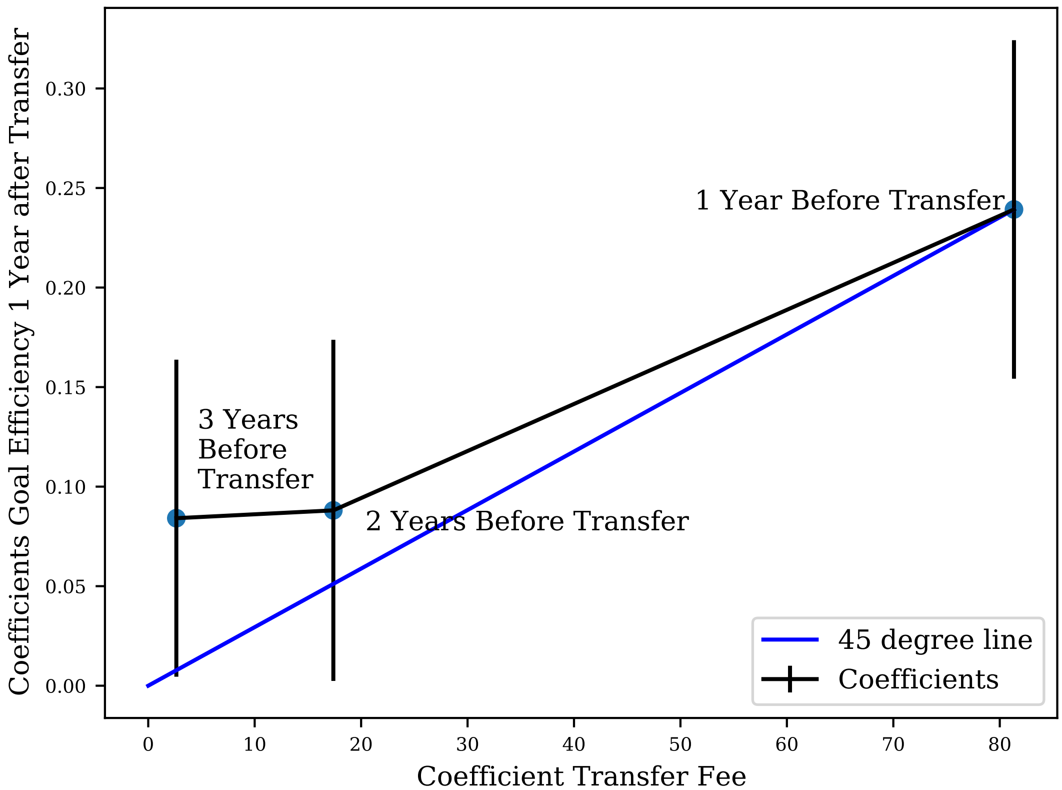


Note: error bars shows interval of 95% significance.

However, the figures show that the plot deviates from 45 degree line, especially so for Figure 6, which means that teams seem to weigh recent performance too heavily. Nevertheless, error bars, which are made using 95% significance level using standard errors for coefficients on goal efficiency after transfer, show that the deviation from the 45 degree line is not statistically significant. Still, the deviation shows that it is possible that teams overweight recent performance, so I compare the ratio of goal efficiency the year before transfer divided by that in earlier years. In the table, Ratio 1 represents $\frac{\beta_{(i,-2)}}{\beta_{(i,-1)}}$, which is the ratio of coefficients of goal efficiency 2 years before transfer and that 1 year before; Ratio 2 represents $\frac{\beta_{(i,-3)}}{\beta_{(i,-1)}}$, which is the ratio of coefficients of goal efficiency 3 years before transfer and that 1 year before (Table 11).

The ratios along with the p-values for test of equality are shown in Table 11. Ratio 1 is consistent between the effect on transfer fee and on performance after transfer, so teams tend to correctly weigh the relative importance of performance one and two years before transfer. On the other hand, the difference in terms of economic magnitude is especially large with Ratio 2 ($\frac{\beta_{(i,-3)}}{\beta_{(i,-1)}}$) as the effect on goal efficiency is 4 to 10 times larger than that on transfer fee, which shows that teams tend to largely under-weigh performance 3 years before

Figure 6: Test of Recency Bias Using Relative Weights of Coefficients of Past Performance from One Year After Transfer



Note: error bars shows interval of 95% significance.

Table 11: Comparison of Performance Ratios

	Effect on log(Transfer+1)	Effect on Goal Efficiency Year 0	Effect on Goal Efficiency Year 1	p Value for Test of Equality Transfer and Year 0 Goal Efficiency	p Value for Test of Equality Transfer and Year 1 Goal Efficiency
Ratio 1	0.214	0.182	0.368	0.539	0.323
	(0.219)	(0.244)	(0.256)		
Ratio 2	0.032	0.139	0.351	0.281	0.118
	(0.115)	(0.144)	(0.243)		

Note: Standard errors are in parenthesis. Bootstrapping methods are used to generate standard errors using standard deviation of the ratio of coefficients. p values are generated using upper tail T-test.

transfer compared to that 1 year before when they pay the transfer fee.

However, despite the large difference in terms of magnitude, the p-values show that none of the test for equality are statistically significant. Even in the most extraordinary case where Ratio 2 for goal efficiency in year 1 is ten times as large as that for transfer fee, the p-value is not even statistically significant at 10% level. If I use the variance and covariance of the coefficients instead of the standard deviation of the ratio of coefficients to compute the standard errors, the standard errors were only slightly smaller and did not influence most of the p-values (Table B.1). The only difference is that the p-value for Ratio 2 between goal efficiency year 1 and transfer fee becomes statistically significant at 10%, which is quite high considering that the former is much larger than the latter.

When comparing Table 11 with Figure 5 and Figure 6 as references, the statistical insignificance becomes understandable. In the figures, most of the deviation of the coefficients from the 45 degree line are not statistically significant different from the 45 degree line. Hence, when analyzing the ratio of coefficients, the ratios are also unlikely to be statistically different. Hence, although there is evidence of teams under-weighing performance 3 years before transfer relative to performance 1 year before transfer, the noise in the data set generates a large standard error, which causes the inequality test between the Ratios in Table 11 to be statistically insignificant.

5 Discussions

5.1 Links to Inefficiency and Irrationality

An argument arising from the result is that whether the findings of sunk cost fallacy may be classified as soccer teams' being inefficient or irrational. On one hand, if rationality and efficiency are defined as maximizing teams' performance, then committing to sunk cost fallacy would be irrational since the teams are not maximizing performance by playing players according to their performance. On the other hand, it is also possible that teams who pay high transfer fees for a player have high expectations in the player's ability. As a result, the teams may need a few seasons of observation of performance to update their Bayesian priors regarding the players' ability; the process of updating Bayesian priors takes time, and therefore there is no irrationality involved (Hinton & Sun, 2020).

The argument is plausible; in fact, it potentially explains one channel through which transfer fee correlates with more playing time, which is that a team may be willing to give a player more time through that higher transfer fee leads to higher hope for the player to perform well in the future. However, such behavior may be classified as inefficient if the purpose of the team is to maximize the team's performance on the field since it is not making choices based on performance. Moreover, such behavior is possible for the first few games after players transfer, but the positive correlation persists throughout several seasons, so the teams must have maintained and utilized prior expectations, due to transfer fee, for a long time. The key to arguing for irrationality is, then, to determine a reasonable length of time during which the teams should decide on playing time independent of initial transfer fee. Such determination is beyond the scope of the paper, but a team should be able to do so within a short period of time since the teams also observe players during practices, which occur frequently. In this case, the coefficients being significantly positive for an entire season or several seasons mean that the teams are likely being irrational with transfer fee.

In contrast, whether teams may be classified as being irrational or inefficiency using the detection of recency bias is unclear. The main difference between the commitment of sunk cost fallacy and that of recency bias is that the former is entirely dependent on teams' choices while the latter depends on negotiation between the buying and the selling team (since all free transfers are excluded from recency bias analysis). Therefore, it may be that teams that buy players are willing to be observant of performance in the more distant past, but the selling teams focus mostly on recent performance (i.e. the performance in the season immediately prior to transfer), which causes the final price negotiated to correlate more strongly with more recent performance.

Hence, in this case, it is difficult to argue that teams thatbuy players in general commit to recency bias. However, the soccer market for players as whole do commit to recency bias. As a result, both buying teams and selling teams lose on opportunity to be more efficient. Assuming that teams would like to purchase players according to their performance after transfer, then the price would be too low for players who played relatively better in the more distant past, which is a loss to the selling team. On the other hand, if a player plays poorly three years ago relative to his performance in the season immediate before transfer, then the price a team pays for this player is likely too high, which is a loss to the purchasing team. Hence, committing to recency bias causes all teams to buy or sell at prices that are less ideal, which means that the entire market may be inefficient due to recency bias.

Nevertheless, it is difficult to attribute the source of the inefficiency to irrationality of the buying or the selling team since we do not know which side has greater bargaining power. In the case that the buying and selling of players are considered as barter outcomes of teams trading players and money, then the negotiation between the buying and selling team determines the final price on the contract curve (the set of Pareto efficienct prices where both buying and selling teams find acceptable). In general, the party with greater bargaining power has a stronger ability to determine the final price on the contract curve, so it is likely that the side with greater bargaining power causes the commitment of recency bias in the market. However, it could be that both the buying and the selling team commit to recency bias, which causes the entire sets of prices on the contract curve to be biased. Hence, a potential topic for future exploration is to identify the party (or parties) responsible for the commitment of recency bias during negotiations of player prices.

5.2 Control Variables

While the significant positive correlation of sunk cost fallacy is highly consistent with numerous prior researches done on sports economics (Camerer & Weber, 1998; Hinton & Sun, 2020; Keefer, 2015, 2017, 2019; Staw & Hoang, 1995), it is the opposite from the most relevant research done by Hackinger (2019), who finds no evidence supporting sunk cost fallacy among German soccer teams. While the data of the papers are different (this paper contains a longer time span and more leagues), the main reason for the difference in coefficients is likely the controls being used. This paper chooses to use within game statistics while Hackinger (2019) used player grades given by magazines. Some people may argue that the grades are better indication of performance for players since they capture qualities of players unobserved by statistics. However, it is undeniable that grades, especially seasonal ones, are based mostly on overall performance instead of efficiency. Overall performance of the player, then, have high collinearity with the dependent variable, playing time. A player who plays more likely has more goals and assists, which results in higher grades, but their efficiency may not be necessarily higher.

Both grades and in-game statistics are performance indices intended to control for performance of players in order to reduce omitted variable bias. The strong contrast between the results of the research papers again indicate the impact control variables may have. The contrast again resonates with the comments of Hinton and Sun (2020) regarding the importance of control in studies for sunk cost fallacy. It is up to interpretation whether the team should base decisions on grades or efficiency. The issue with the former is collinearity with playing time and thus inability to identify players who have performed well but played little time. While the latter rates player in a way that their playing time does not matter, it is possible that the method is unfair to players who have played more. It is likely that a team plays their best players against stronger components and substitute the weak players when the game is intense. Therefore, those with higher ability tend to have lower efficiency than they deserve. It is, then, more ideal to develop a model with respect to within game statistics account for the opponents' strength. It would be a possible extension to weight the efficiency to obtain a fairer control. With sufficient data, an even more ideal option is to include other positions (midfield, defenders, goalies) as part of the study with separate controls for each.

The issue with control variables may be more severe with the analysis of recency bias in this paper. Most successful research that found statistically significant evidence of recency bias used a combination of many in-game statistics to develop a model to generate performance variables (Healy, 2008; Hochberg, 2011). For soccer, however, there is less known ways to evaluate performance systematically, so I only used goal efficiency, arguably the most important statistic for forward, as performance variable. However, only using one variable is not comprehensive and generates a lot of noise as teams likely only evaluate performance partially on goal efficiency. Hence, although the magnitude of the difference of ratios between past performance and more recent performance on performance after transfer versus transfer fee is large, the noise causes the result to be mostly statistically insignificant. Hence, to obtain more compelling and comprehensive evaluation of recency bias, there is need to evaluate players using more compehensive performance evaluations; this performance evaluation likely needs to be different across distinct positions. Gaviao et al. (2020) has generated promising results in evaluating player performance; they used more comprehensive within-game statistics from WhoScored (a website that provides player ratings and statistics for every match recorded) to rank players using Moneyball concept. Unfortunately, the statistics from WhoScored has been inaccessible for this paper.

5.3 Causality for Transfer Fee Sunk Cost in Increasing Playing Time

The causality of higher transfer fee on more playing time is also still dubious at this stage. While there is a positive statistically significant correlation between transfer fee and playing time through regression of difference of variables, using selling team balance as instrumental variable for two stage least squares regression did not generate results that support for causation. Hence, the channel through which transfer fee affects playing time is still unknown. Indeed, it may not necessarily be the case that it is the monetary value of the transfer fee that causes playing time to increase.

It is possible that higher transfer fee means a more difficult negotiation, meaning that the player was harder to obtain. A transfer fee in the millions would involve many hours of team meetings and meetings with other clubs while a free transfer by moving a player from the youth team is very easy to follow through. Therefore, perhaps it is the procedure that accompanies the transfer fee, instead of the fee itself, the varies with the playing time. Such a variation would still be classified as sunk cost fallacy since the time and effort put into the transfer is also a sunk cost, but it is not only the monetary value that make the team to play the players more. Furthermore, as argued in section 5.1, higher transfer fee may mean the higher hope a team places on a player to perform well on a team. In this case, it could be the heuristics of the manager or the head coach in their faith of the players that result in the increase of playing time. In summary, while the robustness checks and the coefficient stability test do a better job at capturing the correlation between he variables, it still does not exclude all extraneous variables and does not enable us to argue for causality. There is potential, therefore, to conduct further studies to look for causality of transfer fee on playing time through the above channels suggested.

6 Conclusions

Overall, the paper finds evidence of sunk cost fallacy and recency bias using transfer fees as fixed cost among teams in the Top 5 European soccer leagues. There is a positive and statistically significant correlation between transfer fee and playing time for the first two or three years after a player transfers to a new team. I eliminate the issue of skewed data and unobserved variable bias through robustness checks and coefficient stability tests. Using ratios between the impact of performance in the more distant past divided by that in the year immediately prior to transfer on performance after transfer and on transfer fee, the paper finds large but statistically insignificant results that suggest soccer teams may commit to recency bias when they purchase players.

The finding for sunk cost fallacy is consistent with previous results that found positive correlation of sunk cost with playing time. The trend that economic magnitude of sunk cost decreases over time is also observed. That sunk cost fallacy may indicate inefficiency suggests that teams may perform better if they can adjust expectations of players' ability independent of transfer fee more quickly. However, the result is opposite to Hackinger (2019), who found no evidence of sunk fallacy using different controls. On the other hand, evidence of recency bias suggests that both buying and selling teams may be better off when the market stops committing to recency bias.

The contrast in the results for sunk cost fallacy and the statistically insignificant result for recency bias analysis encourage further studies to obtain a better model to control for performance—possibility extending to other positions—and to use alternate method to argue for causality of sunk cost in increasing playing time.

References

Altonji, J. G., Elder, T. E., Taber, C., & Zakrajšek, E. (2000). Selection on observed and unobserved variables. [electronic resource]: Assessing the effectiveness of catholic schools. *National Bureau of Economic Research*.

Arkes, H. R., & Blumber, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Performance*, (1), 124.

Borland, J., Lee, L., & Macdonald, R. D. (2011). Escalation effects and the player draft in the afl.". *Labour Economics*, *18*(3), 371–380. https://doi.org/10.1016/j.labeco.2010.12.004

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

Camerer, C. F., & Weber, R. A. (1998). *The econometrics and behavioral economics of escalation of commitment: A re-examination of staw and hoang's nba data*.

Fox, C. (2015). Nba contracts and recency bias: An investigation into irrationality in performance pay markets.

Gaviao, L. O., Sant'Anna, A. P., Lima, G. B. A., & Garcia, P. A. d. A. (2020). Evaluation of soccer players under the moneyball concept. *Journal of Sports Sciences*, *38*(11-12), 1221–1247. https://doi.org/10.1080/02640414.2019.1702280

Hackinger, J. (2019). Ignoring millions of euros: Transfer fees and sunk costs in professional football. *Journal of Economic Psychology*, *75*(Part B). https://doi.org/10.1016/j.joep.2018.10.006

Healy, A. (2008). Do firms have short memories? evidence from major league baseball. *Journal of Sports Economics*, *9*, 407–424.

Hinton, A., & Sun, Y. (2020). The sunk-cost fallacy in the national basketball association: Evidence using player salary and playing time. *Empirical Economics: Journal of the Institute for Advanced Studies, Vienna, Austria*, *59*(2), 1019. https://doi.org/10.1007/s00181-019-01641-4

Hochberg, D. (2011). *The Effect of Contract Year Performance on Free Agent Salary in Major League Baseball*.

Keefer, Q. A. W. (2015). Performance feedback does not eliminate the sunk-cost fallacy: Evidence from professional football. *Journal of Labor Research*, *36*(4), 409–426. https://doi.org/10.1007/s12122-015-9215-y

Keefer, Q. A. W. (2017). The sunk-cost fallacy in the national football league: Salary cap value and playing time. *Journal of Sports Economics*, *18*(3), 282–297.

Keefer, Q. A. W. (2019). Do sunk costs affect expert decision making? evidence from the within-game usage of nfl running backs. *Empirical Economics: Journal of the Institute for Advanced Studies, Vienna, Austria*, *56*(5), 1769. https://doi.org/10.1007/s00181-017-1399-y

Leeds, D., Leeds, M. A., & Motomura, A. (2013). Are sunk costs irrelevant? evidence from playing time in the national basketball association. *Institute of Labor Economics (IZA)*.

Mourao, P. R. (2016). Soccer transfers, team efficiency and the sports cycle in the most valued european soccer leagues – have european soccer teams been efficient in trading players? *Applied Economics*, *48*(56), 5513–5524. https://doi.org/10.1080/00036846.2016.1178851

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *13*(2), 187–204.

Staw, B. M., & Hoang, H. (1995). Sunk costs in the nba: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly*, *40*(3), 474–494.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, (1), 39.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Appendices

A Checking Exogeneity of Instruments with Performance Variables

This section checks the correlation between instrument used in Model 3 and Model 4, which is balance of the selling team during the season of transfer, with performance variables after transfer. Insignificant correlation does not guarantee exogeneity, but demonstrates that the instrument unlikely correlates with player quality.

Table A.1: Effect of Selling Team Numbers of Arrival on Goal Efficiency

	Goal Efficiency 0 Year From Transfer	Goal Efficiency 1 Year From Transfer	Goal Efficiency 2 Year From Transfer	Goal Efficiency 3 Year From Transfer	Goal Efficiency 4 Year From Transfer	Goal Efficiency 5 Year From Transfer
	(1)	(2)	(3)	(4)	(5)	(6)
log(Selling Team Balance + 1)	$-5.94*10^{-6}$	$-3.19*10^{-6}$	$-7.63*10^{-6}$	$-22.07*10^{-6}$	$18.58*10^{-6}$	$35.80*10^{-6}$
	$(3.96*10^{-6})$	$(5.00*10^{-6})$	$(11.71*10^{-6})$	$(21.22*10^{-6})$	$(15.54*10^{-6})$	$(37.09*10^{-6})$
Age	$36.77*10^{-6**}$	$-69.22*10^{-6***}$	$-42.14*10^{-6}$	$100.59*10^{-6}$	$-41.73*10^{-6}$	$124.95*10^{-6}$
	$(18.08*10^{-6})$	$(23.12*10^{-6})$	$(55.12*10^{-6})$	$(98.42*10^{-6})$	$(74.39*10^{-6})$	$(188.35*10^{-6})$
Observations	3,324	1,666	915	486	221	107
R^2	0.163	0.206	0.367	0.235	0.546	0.700
Note:						*p<0.1; **p<0.05; ***p<0.01

Table A.2: Effect of Selling Team Numbers of Arrival on Goal Efficiency

	Points per Appearance 0 Year From Transfer	Points per Appearance 1 Year From Transfer	Points per Appearance 2 Year From Transfer	Points per Appearance 3 Year From Transfer	Points per Appearance 4 Year From Transfer	Points per Appearance 5 Year From Transfer
	(1)	(2)	(3)	(4)	(5)	(6)
log(Selling Team Balance + 1)	$-8.010*10^{-4}$	$-4.158*10^{-4}$	$-15.462*10^{-4}$	$1.925*10^{-4}$	$-9.111*10^{-4}$	$7.155*10^{-4}$
	$(5.641*10^{-4})$	$(7.686*10^{-4})$	$(11.790*10^{-4})$	$(18.168*10^{-4})$	$(28.482*10^{-4})$	$(49.350*10^{-4})$
Age	$7.982*10^{-3***}$	$-3.578*10^{-3}$	$-6.147*10^{-3}$	$-3.000*10^{-3}$	$-15.399*10^{-3}$	$19.611*10^{-3}$
	$(2.576*10^{-3})$	$(3.554*10^{-3})$	$(5.552*10^{-3})$	$(8.427*10^{-3})$	$(13.633*10^{-3})$	$(25.063*10^{-3})$
Observations	3,324	1,666	915	486	221	107
R^2	0.331	0.391	0.440	0.533	0.669	0.832
Note:						*p<0.1; **p<0.05; ***p<0.01

From Table A.1 and Table A.2, balance has insignificant correlation with performance after transfer, meaning that the instrument does not indicate player quality.

B Alternate Standard Errors for Recency Bias Ratio Comparison

In addition to using the standard deviation from bootstrapping as the standard errors for ratios in Table 11, we can obtain the standard deviation of the ratios from the variance and covariance of the coefficients from bootstrapping (Healy, 2008). The advantage in this alternate method to obtain the standard error is to avoid occasions when coefficient on goal efficiency in the season immediate before transfer (i.e. the coefficient on $goal_{(i,-1)}$) is small, which may result in the standard error of the ratios to be too large.

By comparing Table 11 and Table B.1, we see the standard errors are indeed smaller when using variance and covariance. However, the decrease in standard error is small, and the only difference is that Ratio 2 for equality between transfer fee and year 1 goal efficiency is now at 10% significance. Hence, using standard deviation of ratio from bootstrapping directly is accurate in general.

Table B.1: Comparison of Performance Ratios Using Alternate Standard Errors

	Effect on log(Transfer+1)	Effect on Goal Efficiency Year 0	Effect on Goal Efficiency Year 1	p Value for Test of Equality Transfer and Year 0 Goal Efficiency	p Value for Test of Equality Transfer and Year 1 Goal Efficiency
Ratio 1	0.214	0.182	0.368	0.544	0.300
	(0.186)	(0.223)	(0.228)		
Ratio 2	0.032	0.139	0.351	0.266	0.090
	(0.106)	(0.135)	(0.213)		

Note: Standard errors are in parenthesis. Bootstrapping methods are used to generate standard errors using variance and covariance of coefficients. p values are generated using upper tail T-test.