DOES BETTER SPORTS PERFORMANCE GENERATE HIGHER REVENUES IN THE ENGLISH PREMIER LEAGUE? A PANEL DATA APPROACH

Marina Schloesser[1]✉, Václav Adamec[1]

[1] *Mendel University in Brno, Czech Republic*

EUROPEAN JOURNAL OF BUSINESS SCIENCE AND TECHNOLOGY

Volume 9 Issue 1
ISSN 2694-7161
www.ejobsat.com

ABSTRACT

In this paper, we examined the relationship of sports performance and revenue generation in the English Premier League (EPL) to understand how performance on the field impacts financial performance of professional football clubs. Further, we verified if increased wage expenses help improve sports performance. Independent dynamic models were estimated by GMM on panel data including $N = 28$ EPL teams and on a reduced data set excluding the top six teams ($N = 22$), spanning from the 2008/2009 to 2018/2019 seasons ($T = 11$). The results of the GMM models confirmed that sports performance and revenue generation significantly correlate. Teams with better sports performance do generate higher revenues. Additionally, higher wage expenses result in better sports performance. A positive relationship of the variables in both hypotheses were established in both directions (full data). In all analyses of reduced data, the parameters of interest are nonsignificant. Dependencies exist due to the top teams.

KEY WORDS

revenue, sports performance, panel data, Generalized Method of Moments, wage expenses, football

JEL CODES

C23, D22, J30, Z23

1 INTRODUCTION

Over the last decades, professional football has turned into a multi-million-dollar global business operation with fans following from all over the world. This in turn led to a world-wide competition for fan attention and money. In an increasingly connected world, professional football teams not only need to compete against domestic league competitors but also against other global football leagues and even a variety of other sports across the world in order to maximize revenues and fan engagement.

SCHLOESSER, Marina, and ADAMEC, Václav. 2023. Does Better Sports Performance Generate Higher Revenues in the English Premier League? A Panel Data Approach. *European Journal of Business Science and Technology*, 9 (1): 21–36. ISSN 2694-7161, DOI 10.11118/ejobsat.2023.006.

Whilst debate regarding the key objective of professional football clubs – to be either revenue maximizers or utility maximizers – has been going on for decades (see for example Sloane, 1971), the continuous revenue increase amongst football clubs in Europe, particularly over the past two decades, suggests that the economic aspect is the most important one for professional sports organizations.

Previous studies showed that by improving sports performance, club revenues can be increased. Szymanski (1998) reported that league performance and club revenues are correlated. More specifically, he claimed that "better league performance leads to higher revenues that occur as a result of increased attendance, higher ticket prices, increased level of sponsorship, and income from merchandising and TV rights."

Since professional sport has undergone a variety of changes and innovations over the last years, the objective of this paper is to identify and quantify the relationship of revenue generation and sports performance in the English Premier League (EPL) on data collected between 2008/2009 and 2018/2019. This research will confirm or reject previous academic research undertaken on different professional football leagues. Also, we propose to identify the relationship between sports performance and wage expenditures.

The added value of current approach is use of a larger data set, as previous research data apart from Szymanski (1998) used data beyond ten seasons. Further, we applied a novel statistical approach to more recent and larger empirical data to include more teams than past research. The projected panel data analysis is expected to confirm possible relationships between sports performance and revenues, and also between wage expenses and sports performance to illustrate full cycle of potential revenue generation in professional European Football.

The rest of the paper is structured as follows: section two presents an overview of the most relevant literature on this topic. This provides background on the topic and help understand the status of academic research. Section three describes the data and methodology applied in this research. Results are presented and discussed in section four. The paper concludes in section five comprising a summary of the results, research questions and identification of potential strengths and weaknesses of the presented approach.

2 THEORETICAL FRAMEWORK

With football becoming broadly popular and relevant across the world, interest from an academic point of view has visibly grown. The wealth of data currently measured, tracked and publicly available enabled the possibilities for endless research. Several authors have analyzed the relationship of sports performance and financial performance of professional football clubs in several countries and leagues.

The first key paper was published by Szymanski (1998) who collected data from 69 clubs in the English Premier League over a hundred years and analyzed the relationship of league position and revenue, as well as league position and revenue. He concluded that league performance did not bear an impact on revenues. However, when the relationship of revenue and expenses to league performance were examined, the results showed much more useful outcomes and lead Szymanski (1998) to formulate two general concepts: 1. Better league performance leads to higher revenues and 2. Increased wage expenditure leads to better league performance.

Pinnuck and Potter (2006) took a similar approach on the Australian Football League (AFL) looking at the relationship of sports performance and different revenue streams. The authors focused on factors that influenced financial performance of the AFL over ten seasons. Pinnuck and Potter came to several insightful findings. Firstly, match attendance and sports success are strongly related. And secondly, membership figures are positively

Tab. 1: Previous empirical studies analyzing sport performance, revenues, and efficiency

Author	Country	Variables	Key results
Szymanski (1998)	England	League position and revenues and wage expenses and league position	Better league position leads to higher revenues; increased wage expenditure leads to better league performance
Gerrard (2005)	England	Sports performance, profitability, wage costs, playing quality, revenues, fan base	Team revenue is positively related to sports performance. Operating margin is negatively related to sports performance
Pinnuck and Potter (2006)	Australia	Match attendance, marketing, membership revenues	Sports performance has a positive impact on attendance and marketing revenues
Haas (2003)	England	Wages/salaries, points, total revenue	Sport ranking is not significantly related to the efficiency ranking
Guzmán and Morrow (2007)	England	Staff costs, points, total revenue	Efficiency scores are not correlated with sports ranking
Ribeiro and Lima (2012)	Portugal	Wages, league ranking	Efficiency rank and league rank don't correlate

influenced by past sports success and marketing expenses.

In his paper, "a resource-utilization model of organizational efficiency in professional sports teams," Gerrard (2005) analyzed whether revenues and sports performance relate. He affirmed that they positively relate and estimated that for every 1% improvement in points accumulated per season, revenues increase by 0.81% relative to the league's average. The authors also reported that financial performance is negatively related to sport performance, as with every 1% increase in sports performance, operating margins decrease by 0.25%.

Rather than exploring pure financial performance vs. sports performance, Haas (2003) investigated efficiency scores. In the productive efficiency, Haas researched the actual team performance in the English Premier League vs. the possible performance outcome, given the investments in talents made by the respective club in season 2000/2001. Through his Data Envelopment Analysis (DEA) approach, he came to the conclusion that the league ranking of the clubs at the end of the season are not related to the ranking based on the efficiency scores. More specifically, clubs with the highest investments in players and coaches measured by wage expenses are not achieving the success their investments would suggest.

Guzmán and Morrow (2007) increased the sample size to six seasons to explore the relationship of efficiency scores and sports performance in the EPL between seasons 1997/98 and 2002/03. The longer investigation period does not impact the findings as results are in line with Haas (2003) indicating no significant relationship between efficiency and sports performance. It is interesting, however, that Guzmán and Morrow seemed to detect an inverse relationship of the two, as the lower ranked teams showed better efficiency scores than most of the top ranked teams.

Later, Ribeiro and Lima (2012) expanded the investigated timeline to seven seasons covering the period from 2002/03 to 2008/09. They also applied the DEA method but using the Portuguese Football League as their data source for research. The authors found a negative relationship between sports performance and efficiency scores: clubs which perform well in the league ranking and employ the best players often give poor efficiency scores, while smaller clubs with less expenses on players seem to extract more out of their resources and achieve higher efficiency scores. Tab. 1 summarizes the most relevant historical studies and respective key findings.

As the literature reveals, there is great interest in football research. Professional football is not only increasing in terms of global reach, popularity, and revenues. With the data now accessible, there is heightened interest in the impact of sports performance on revenues. Our recent data set enables us to examine whether these relationships hold or have changed in the recent years. By applying an alternative approach of panel data analysis, we seek to uphold past academic papers and also illustrate the full cycle of revenue increase.

3 DATA AND METHODS

European football continues to enjoy its success when it comes to revenue growth, with the largest share of revenue coming from the "big five" leagues, i.e., England, Germany, France, Spain and Italy. The EPL generated the highest absolute growth in 2018/19, and "continues to generate the highest revenues across the 'big five'" (€ 5.9 billion). The 20 EPL clubs' combined revenue grew by 7% (Deloitte, 2020). We chose the EPL for present research, as the league not only excels in revenue generation amongst all football leagues globally, but since comprehensive and reputable data sources are available. Whilst past researchers also focused on the English Premier League, we see the existing research on the EPL as advantageous, as it presents greater opportunities for comparison and for identifying changes over time.

The data used in this research covers information on football teams competing in the EPL from season 2008/09 to 2018/19. Each season has 20 teams competing in the EPL to become the league's champion. The champion is the team that collected the most points during the season where three points are awarded for a win, one point for a draw and no point for a defeat. As per rules and regulations of the EPL, each season the teams holding the three lowest rankings of the league's table (rank 18, 19 and 20), i.e., the teams with the lowest number of points, are demoted to the EFL Championship – until 2016 known as the Football League Championship – whilst the two top teams of the EFL Championship plus one additional team are promoted to the EPL. Therefore, instead of 20 consecutive team observations, we identified 36 different teams over the course of the eleven seasons. Data on team revenues and wage expenses were retrieved from the Deloitte Annual Review of Football Finance from 2010 to 2020. Information on league position and points accumulated at the end of season were gathered from kicker.de (2008–2019), the online platform of the renowned German sports magazine published by the Olympia-Verlag GmbH. The same source was also used in previous academic research (see Haas et al., 2004).

In this paper, we researched the following two hypotheses: Hypothesis 1 (H_1) is formulated "better sports performance leads to higher revenues". For sports performance, we used the total points a team accumulated over each season. While we also collected data on league positions, we decided against its use as the league ranks are also based on the points the team obtained per season. Therefore, total points as an input variable seemed more reasonable. This approach was also undertaken by many researchers looking into this topic, for example Haas (2003), Haas et al. (2004) or Guzmán and Morrow (2007).

Hypothesis 2 (H_2) suggests that "better players will bring better sports performance". The response variable is cumulated points end of season; the independent variable for Hypothesis 2 is wage expenses (GBP mil). A panel data analysis was chosen as "panel data have space as well as time dimensions" (Gujarati et al., 2012), which makes them suitable for this type of research. It allows not only to explore change and impact over time but also to identify top performing teams. The very poor performing teams eliminated themselves from the analysis by being relegated at the end of the season.

3.1 Data Samples

EPL relegation policies cause our data set to be an unbalanced panel as not all teams can stay in the first league every season and consequently the teams in our dataset do not have the same number of observations. Moreover, we excluded teams that had two or fewer observations as their influence over the eleven seasons is small and do not add useful information to the Generalized Methods of Moments (GMM) estimation of the dynamic models of panel data. Specifically, Birmingham City, Blackpool, Brighton & Hove Albion, City Cardiff City, Huddersfield Town, Middlesbrough, Portsmouth and Reading FC were excluded from the analysis, for this reason. 28 teams remained thus in the dataset, designated as full data.

The time dimension in the current data is season from 2008/2009 until 2018/2019. After the 2019 season, we discontinued the data, as COVID-19 pandemics strongly affected revenue generation in this sport sector. For instance, fans were not allowed to be at the game, or were allowed in limited numbers in the stadium thereby negatively affecting matchday revenues. In our data, teams are cross-sectional units, namely the $N = 28$ teams that competed in the EPL for at least three seasons between 2008/09 and 2018/19 ($T = 3$ to 11). The total number of available observations in the full data is $n = 205$.

Furthermore, we created a reduced data set by eliminating the most successful teams in terms of revenue generation per season, points accumulated per season as well as wage expenses per season. There seem to be sizeable differences in level and variation among the top EPL teams and the remaining ones. Consequently, Arsenal FC, FC Chelsea FC, Liverpool FC, Manchester City FC, Manchester United FC and Tottenham Hotspur FC were removed from the full data set. This was done primarily to explore robustness of the econometric analyses. Consequently, for H_1 and H_2 we estimated the GMM system and difference models first on the full dataset, followed by estimation on the reduced data. The number of observations in the reduced data is $n = 139$ ($N = 22$).

3.2 Data Characteristics and Econometric Models

According to Greene (2002) the analysis of longitudinal data is subject of one of the most active and innovative bodies of literature in econometrics. Gujarati et al. (2012) state that "by combining time series of cross-section observations, panel data gives more informative data, more variability, less collinearity among variables, more degrees of freedom and more efficiency". For these reasons, a panel data analysis was chosen in the current research.

To estimate panel data models, we used the R program, version 4.1.2. R is a language and environment for statistical computing (R Core Team, 2021). Further, we applied the R extensible package plm, version 2.4-3. suitable to estimate numerous linear models of panel data and make robust inferences (Croissant and Millo, 2018).

To investigate the impact of sports performance on revenues we used points as the input variable to represent sports performance. Tab. 2 summarizes the variables employed for exploration of hypothesis 1.

Further, we explored the relationship of wage expenses and sports performance as previously carried out by other authors, for example see Szymanski (1998) or Gerrard (2005). We aim to verify the theory of Szymanski (1998) that "better players win more matches", i.e., wage expenses and sports performance positively correlate. Tab. 3 provides the variables used in the panel data models to verify hypothesis 2.

Fig. 1 and 2 illustrate the level and variation of revenues and points accumulated by the teams in the EPL. Fig. 3 depicts the identical information for player wages.

Generally, there are several methods of analysing longitudinal data. Due to probable dynamic nature of the relationships between the researched variables, we specified and estimated two different dynamic models of panel data: the system GMM model of Blundell and Bond (1998) and the difference GMM model described by Arellano and Bond (1991). Estimation technique for both models was the Generalized Method of Moments (GMM), chosen

Tab. 2: Variables used for verification of hypothesis 1

Category	**Variables**	**Type of variable**	**Unit**	**Source**
Revenues	Total revenues	Dependent variable	GBP mil.	Deloitte Annual Review of Football Finance
Sports performance	Points end of season	Independent variable	Points	www.kicker.de

Tab. 3: Variables used for verification of hypothesis 2

Category	**Variables**	**Type of variable**	**Unit**	**Source**
Sports performance	Points end of season	Dependent variable	Points	www.kicker.de
Wage costs	Wage expenses	Independent variable	GBP mil.	Deloitte Annual Review of Football Finance

primarily to handle likely correlation between the explanatory variable and the current or lagged error term, so called endogeneity issue leading to inconsistent estimates with some traditional estimation approaches. The GMM method applies lagged values of the dependent variable as instrumental variables to address the endogeneity problem.

Consider an initial model of panel data with lagged dependent variable, current and lagged independent variable and time dummies on the right-hand side of the model equation, as shown in (1).

$$y_{it} = c + \phi_1\, y_{it-1} + \phi_2\, y_{it-2} + \\ + \beta_1\, x_{it} + \beta_2\, x_{it-1} + \\ + \sum_{t=2}^{T} \gamma_t\, D_t + (\alpha_i + \varepsilon_{it}), \tag{1}$$

where c is the intercept, ϕ_1 and ϕ_2 are the parameters for the lagged dependent variables, β_1 and β_2 are the coefficients for the current

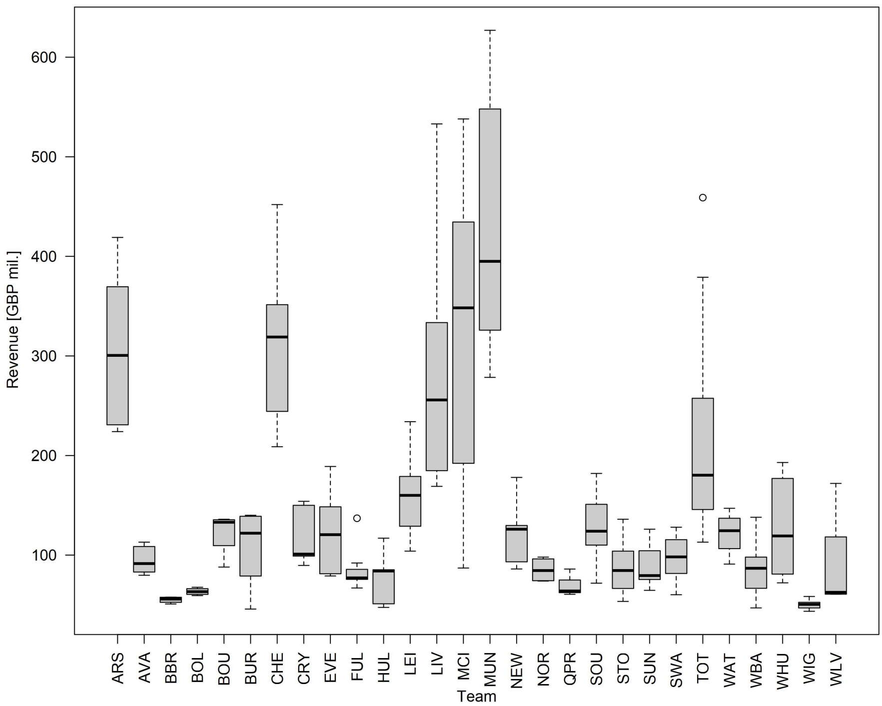

Fig. 1: Boxplots of Revenue (GBP mil.) by EPL team

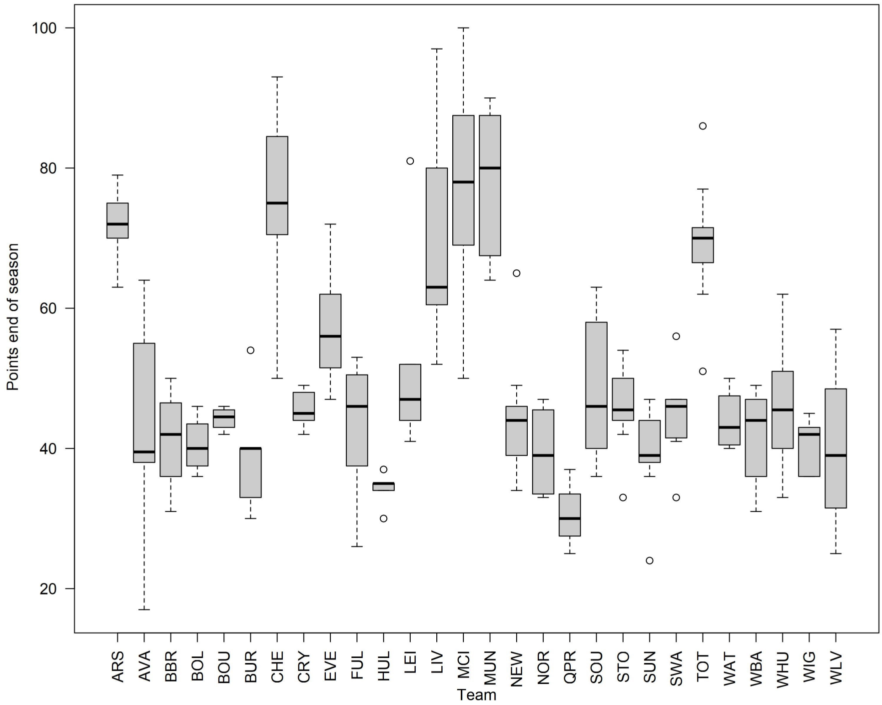

Fig. 2: Boxplots of Points end of season by EPL team

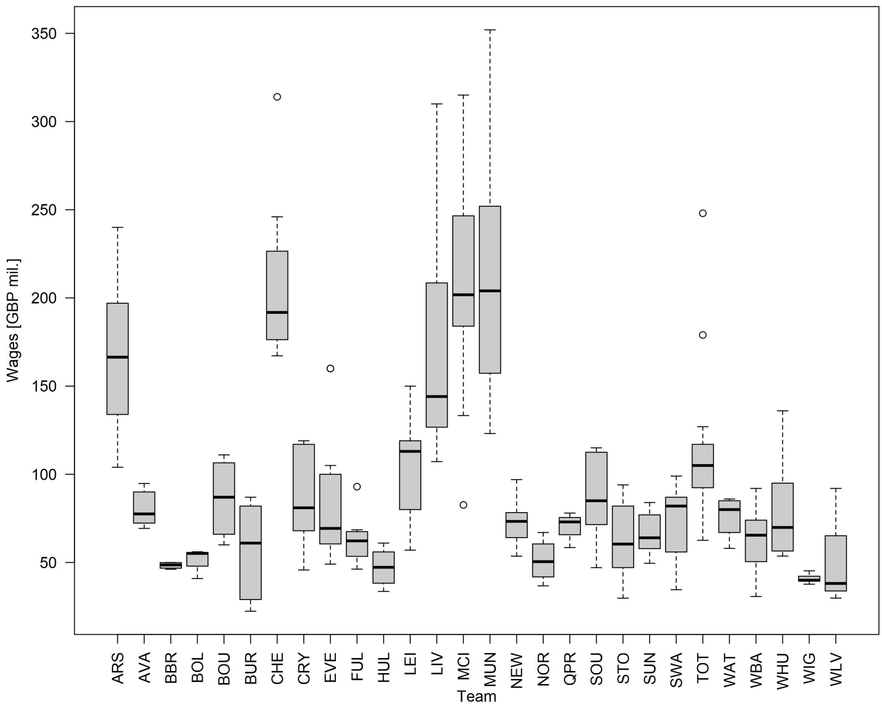

Fig. 3: Boxplots of Wage expenses (GBP mil.) by EPL team

and lagged explanatory variables of interest, γ_t is parameter for the corresponding t-th season dummy, α_i denote the constants associated with the fixed effects (cross-section units), and ε_{it} is the random error term. This model is later transformed to difference GMM model via 1st order differencing $\Delta y_{it} = y_{it} - y_{it-1}$ applied to both sides of equation (1), which removes the individual fixed effects, as shown in equation (2). The method of differencing, however, may possibly increase gaps between observations with unbalanced panel data.

$$\begin{aligned}\Delta y_{it} = \phi_1\, \Delta y_{it-1} + \phi_2\, \Delta y_{it-2} + \\ + \beta_1\, \Delta x_{it} + \beta_2\, \Delta x_{it-1} + \\ + \sum_{t=2}^{T} \gamma_t\, D_t + \Delta\varepsilon_{it}.\end{aligned} \tag{2}$$

The differenced GMM model was complemented with lagged dependent variables, as instruments to handle the issue of endogeneity.

Another variant is the system GMM model by Blundell and Bond (1998). In this model, the dependent variable considered in model (1) is a random walk and its lagged first-order differences serve as model instruments. Estimation of dynamic GMM models on panel data were obtained via the two-step estimation, known to provide better quality estimates. The number of lagged instruments included in GMM dynamic equations was restricted to one for both variants of GMM models. Robust inference was uniformly preferred in GMM model output and statistical tests. Blundell et al. (2000) state that use of the system GMM estimator not only improves the estimator precision but also greatly reduces finite sample bias. GMM estimators are generally applicable when there are independent variables that are not strictly exogenous, i.e., they are correlated with past and possibly current realizations of the error term, with fixed effects, and under heteroskedasticity and (or) autocorrelation within individuals (Roodman, 2009a).

Diagnostics of the estimated GMM models was secured with the Hansen–Sargan test, dubbed J-test (Hansen, 1982; Sargan, 1958) of overidentifying restrictions with weights taken from the two-step estimation procedure. The J-test verifies the null hypothesis of valid instruments. While a too low p-value of J-test may indicate poorly chosen instruments, a p-value close to unity suggests too many instruments, as dealt with in Roodman (2009b). Also, we applied independently the Arellano and Bond (1991) tests of no serial correlation of the error term with tested lags one or two. The Arellano-Bond test variant with robust estimators of the covariance matrix was used.

4 RESULTS AND DISCUSSION

Due to suspected non-stationarity of the variables revenues, wages and sport performance in some panels, we early run independent unit root tests based on LM statistic, as devised by Hadri (2000). Since non-stationarity tests generally show no tolerance to missing values in panels, the Hadri tests with a drift were applied to the six top EPL teams only with significant p-values < 0.001. Occurrence of unit root in at least one panel was thereby confirmed for all variables.

In this chapter, we primarily present the results in form of the econometric models verifying first hypothesis 1: “better sports performance leads to higher revenues” and then hypothesis 2: “better players will bring better sports performance”, examining the relationship of sport performance and wage expenses. Presented GMM models include lagged values of the response variable with lags one and two, time (season) dummies and independent variables from the current and past period with lag one. For hypothesis 2 we applied one lag only to the response variable in the GMM model when testing the reverse direction of the relationship.

4.1 Results for Hypothesis 1: Models of Revenues

In this section, we present GMM models of the current revenues as a function of the revenues lagged two seasons and current points and points from the previous season. The GMM models also include dummies for seasons. Tab. 4 presents the estimated coefficients and significance tests for the system and difference GMM dynamic model of panel data for revenue obtained from full the data. Further, in order to verify the estimated GMM models, we performed a set of diagnostic tests available in the plm library, which are also displayed in Tab. 4.

Tab. 4: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Revenue (mil. GBP) with full data

	Coefficient (std. error)	**Coefficient (std. error)**
Revenues $(t-1)$	0.915***	0.320**
	(0.124)	(0.143)
Revenues $(t-2)$	0.034	0.780***
	(0.139)	(0.179)
Points EoS (t)	0.512***	0.571***
	(0.184)	(0.157)
Points EoS $(t-1)$	−0.619***	1.170***
	(0.234)	(0.336)

Diagnostic test	**Statistic (df)**	***p*-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 8.758$ (17)	0.948
	$\chi^2 = 5.787$ (6)	0.447
Test for 1st order serial correlation	$z = -1.090$	0.276
	$z = -0.371$	0.711
Test for 2nd order serial correlation	$z = 0.774$	0.439
	$z = -1.215$	0.225

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

With the full data we observed positive and significant estimated parameters for the current and lagged points end of season in the difference and in the system model. The p-values of the significance tests were below the 5% significance boundary. The models therefore suggest positive impact of current and past player performance on the current team revenues. The Sargan-Hansen test of overidentifying restrictions shows that the instruments for the above GMM models were selected appropriately, as indicated by the non-significant p-values. The respective diagnostic tests for autocorrelation pointed out that respective error terms were free from serial correlation.

GMM estimates of the identical models on the reduced data as well as the diagnostic tests available in plm library of R, are displayed in Tab. 5 to establish dependency of the results on data sub setting.

Tab. 5: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Revenue (mil. GBP) with reduced data

	Coefficient (std. error)	**Coefficient (std. error)**
Revenues $(t-1)$	0.278	−0.141
	(0.202)	(0.385)
Revenues $(t-2)$	0.197*	−0.068
	(0.105)	(0.497)
Points EoS (t)	0.372*	0.491***
	(0.196)	(0.176)
Points EoS $(t-1)$	0.666	0.679*
	(0.416)	(0.397)

Diagnostic test	**Statistic (df)**	***p*-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 9.440$ (17)	0.926
	$\chi^2 = 2.347$ (6)	0.885
Test for 1st order serial correlation	$z = -0.885$	0.376
	$z = -0.326$	0.745
Test for 2nd order serial correlation	$z = 0.138$	0.890
	$z = 0.266$	0.790

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Estimated coefficients from the dynamic GMM models of the current revenues (reduced data) show significant impact of the current points end of season on the current revenues in the difference model, however significant influence ($\alpha = 0.1$) of the lagged points can be seen in the difference model only. Equality of the parameters obtained from full and reduced data, however, could not be statistically tested. Consequently, we were able to confirm the H_1 hypothesis that player performance, expressed as points increases team revenues in the EPL. The validity of H_1 appears to hold in the tested direction for both full and reduced data, despite existing differences in the strength of the relationship, which seems to be more evident in the full data. The p-value of the Sargan test points out that the model instruments were chosen appropriately. The Arellano-Bond verification tests for serial correlation imply that error terms were not serially correlated in neither the system nor difference GMM model.

To describe the relationship in the opposite direction, primarily to learn, whether current or lagged revenues influence the current sport performance expressed as points end of season, we applied the GMM estimator on the dynamic model specified in the reverse direction. Hence, we modelled the current points end of season as a function of lagged points (lags 1 and 2), and current and lagged revenues. Parameter estimates obtained on full data and output of the diagnostic tests performed to verify the above mentioned GMM models are displayed in Tab. 6.

Tab. 6: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Points end of Season with full data

	Coefficient (std. error)	**Coefficient (std. error)**
Points EoS $(t-1)$	0.136	−0.500**
	(0.159)	(0.217)
Points EoS $(t-2)$	0.105	−0.828***
	(0.144)	(0.311)
Revenues (t)	0.149***	0.058
	(0.410)	(0.051)
Revenues $(t-1)$	−0.066*	0.234***
	(0.034)	(0.079)

Diagnostic test	**Statistic (df)**	**p-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 18.214$ (17)	0.376
	$\chi^2 = 5.333$ (6)	0.502
Test for 1st order serial correlation	$z = -1.739$	0.082
	$z = -1.369$	0.171
Test for 2nd order serial correlation	$z = -0.292$	0.770
	$z = 1.296$	0.195

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Model coefficients obtained from the full data show significant relationships between current revenues and current points in the system GMM model. The difference model thereby shows evidence that lagged revenues exerting a statistically significant impact upon current sport performance (points) in the EPL. The p-values of the Sargan-Hansen J-tests confirm that the model instruments were chosen in a suitable way, as they show no problematic instruments with p-values of acceptable size. The verification tests for autocorrelation first and second order indicate no significant autocorrelations of the residual terms.

Furthermore, we re-estimated the same model to check for robustness on the reduced data, where the top teams have been excluded. GMM estimates and standard diagnostic tests to verify the above GMM models are displayed in Tab. 7.

Tab. 7: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Points end of Season with reduced data

	Coefficient (std. error)	**Coefficient (std. error)**
Points EoS $(t-1)$	0.233	−2.125
	(0.322)	(2.525)
Points EoS $(t-2)$	0.245	−2.806
	(0.339)	(4.296)
Revenues (t)	0.243	0.430*
	(0.403)	(0.221)
Revenues $(t-1)$	−0.146	1.309
	(0.183)	(1.884)

Diagnostic test	**Statistic (df)**	**p-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 11.022$ (17)	0.855
	$\chi^2 = 2.723$ (6)	0.843
Test for 1st order serial correlation	$z = -1.170$	0.242
	$z = 0.177$	0.859
Test for 2nd order serial correlation	$z = -0.870$	0.384
	$z = -0.458$	0.647

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Estimates of the GMM models from the reduced data (see Tab. 7) show no significant results for the variables of interest in the reverse direction (current and lagged revenues affecting the current cumulated points). Although the corresponding p-values are just above the 5% significance boundary, they are small, but not significant. This might imply that it is primarily the top EPL teams that attract the most revenues to stimulate better game performance. The p-values of the Sargan-Hansen tests confirm that model instruments were chosen correctly. The verification tests for autocorrelation indicate no autocorrelation of the first and second order in both GMM models.

Parameter estimates of our GMM models on H_1 confirm that sport performance and revenue generation are significantly correlated; better playing clubs do generate higher revenues. This relationship still holds true for the full data and when the top performing teams were removed. We also found significant results in the reverse direction in the system and difference GMM model, detecting a positive impact of revenues on sport performance. Nonetheless, after removing the top teams the relevant parameters become no longer statistically significant and the relationship does not seem to hold.

4.2 Results for Hypothesis 2: Models of Points End of Season

In hypothesis 2, we are interested in identifying and quantifying a hypothetical relationship between wage expenditures and sport performance. For this reason, we estimated system and difference GMM models on the variables specified for hypothesis 2 on the full data, the reduced data, in the tested as well as the reversed direction of the hypothesized relationships on the full or reduced data.

Tab. 8: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Points end of Season with full data

	Coefficient (std. error)	**Coefficient (std. error)**
Points EoS $(t-1)$	0.058	−0.611***
	(0.165)	(0.205)
Points EoS $(t-2)$	−0.141	−0.553
	(0.160)	(0.345)
Wages (t)	0.123***	0.190***
	(0.040)	(0.054)
Wages $(t-1)$	0.123**	0.207***
	(0.056)	(0.036)

Diagnostic test	**Statistic (df)**	**p-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 17.245$ (17)	0.438
	$\chi^2 = 5.224$ (6)	0.515
Test for 1st order serial correlation	$z = -1.756$	0.079
	$z = -0.729$	0.466
Test for 2nd order serial correlation	$z = -0.268$	0.788
	$z = -0.319$	0.750

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Presently, we modelled the current points end of season as a function of points from the previous season one and two seasons back and then as a function of the current and lagged wages. Coefficients of the system and difference GMM models obtained on the full data are displayed in Tab. 8 along with the diagnostic tests of the plm library to verify the GMM models.

Estimated parameters of the system and difference GMM model from full data show that current and lagged wages do significantly increase cumulated points end of season in the current season. The non-significant p-values of the Sargan-Hansen J-tests indicate that the model instruments appear to be selected in a suitable way. The verification tests for autocorrelation detect that the error terms are free from autocorrelation as shown by the respective p-values above the 5% significance level for both models.

Furthermore, we re-estimated the above-mentioned models on the reduced data to establish robustness of the estimated model parameters towards data segmentation. Coefficient estimates and diagnostic tests can be found in Tab. 9.

Tab. 9: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Points end of Season with reduced data

	Coefficient (std. error)	**Coefficient (std. error)**
Points EoS $(t-1)$	0.217	−0.535
	(0.275)	(0.536)
Points EoS $(t-2)$	0.176	−0.267
	(0.208)	(0.694)
Wages (t)	0.252*	0.238
	(0.146)	(0.166)
Wages $(t-1)$	−0.155	0.121
	(0.237)	(0.310)

Diagnostic test	**Statistic (df)**	**p-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 10.019$ (17)	0.903
	$\chi^2 = 4.098$ (6)	0.663
Test for 1st order serial correlation	$z = -1.092$	0.275
	$z = -0.119$	0.905
Test for 2nd order serial correlation	$z = -1.156$	0.248
	$z = -0.794$	0.427

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

While the p-value for wages in the system model was reasonably small ($\alpha = 0.1$), both

GMM models estimated on the reduced data showed no significant relationships specified in the tested direction of hypothesis 2.

Additionally, we verified the hypothetical dependency specified in hypothesis 2 in the reversed direction, i.e. whether current and lagged points end of season influence current wages. For this purpose, we applied the GMM estimator on dynamic models in the opposite direction early on the full and then on the reduced data. In Tab. 10, we present GMM models estimated on the full data including a single lag of the response variable. Output of the diagnostic tests follows.

Tab. 10: Estimated parameters and significance z-tests for the system GMM model (left) and difference GMM model (right) of Wages (mil. GBP) with full data

	Coefficient (std. error)	**Coefficient (std. error)**
Wages $(t-1)$	0.832***	0.624
	(0.094)	(0.385)
Points EoS (t)	0.265*	0.342*
	(0.153)	(0.175)
Points EoS $(t-1)$	0.527***	0.672***
	(0.157)	(0.242)

Diagnostic test	**Statistic (df)**	***p*-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 13.614$ (19)	0.806
	$\chi^2 = 12.633$ (8)	0.125
Test for 1st order serial correlation	$z = -2.070$	0.038
	$z = -1.957$	0.050
Test for 2nd order serial correlation	$z = 0.812$	0.417
	$z = 0.576$	0.565

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Estimated model coefficients from the full data show statistically significant impact of the lagged cumulative points upon current Wages both in the system and the difference models. Consequently, the GMM models establish that past sport performance, measured as cumulated points per season impact the current wage expenses in the EPL. The non-significant p-values of the Sargan-Hansen J-tests for overidentification confirm appropriateness of the selected model instruments with properly sized p-values. The verification tests for autocorrelation first and second order identified that error terms were free from autocorrelation, except for the 1st order serial dependency test applied to the system GMM model.

As for the previous main hypotheses, we re-estimated the GMM models on the reduced data to check for sensitivity of the model estimates and tests thereof towards data subsampling. Model estimates together with the diagnostic tests are presented in Tab. 11.

Tab. 11: Estimated parameters and standard errors for the system GMM model (left) and difference GMM model (right) of Wages (mil. GBP) with reduced data

	Coefficient (std. error)	**Coefficient (std. error)**
Wages $(t-1)$	0.666***	0.114
	(0.080)	(0.181)
Points EoS (t)	0.200*	0.175
	(0.107)	(0.126)
Points EoS $(t-1)$	0.179	0.315*
	(0.198)	(0.170)

Diagnostic test	**Statistic (df)**	***p*-value**
Sargan-Hansen J-test for overidentifying restrictions	$\chi^2 = 9.209$ (19)	0.970
	$\chi^2 = 10.370$ (8)	0.240
Test for 1st order serial correlation	$z = -2.135$	0.033
	$z = -1.904$	0.057
Test for 2nd order serial correlation	$z = 1.940$	0.053
	$z = 1.730$	0.084

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.

Results of the system GMM model and difference GMM model of Wages (mil. GBP) estimated with reduced data show positive but nonsignificant estimated parameter for the lagged points coefficients. Therefore, nonsignificant relationship between points and wages can be reported. The system GMM model suggests that impact of points upon wages is markedly influenced by presence of the top teams in the EPL. The investigated effect in the reduced data appears to be of lesser magnitude compared to the unrestricted data. The Sargan-Hansen J-tests confirm that current model instruments were properly selected with the non-significant p-value in both GMM models. The verification tests for autocorrelation indicate dependency for the 1st order serial dependency test applied to the system model. Otherwise, no serial dependencies of the first or the second order were detected in either GMM model.

Estimated coefficients of the GMM models on full data affirm that wage expenses and sport performance correlate; better players do bring better sports performance. However, our analyses showed no significant results with reduced data. This might suggest that mostly the top teams can afford to employ the very best players that ultimately make the difference on the field, while the average clubs mostly engage mediocre players being paid on the size of the club's budget.

We also found significant results in the reversed direction in both the system and difference GMM models, detecting a positive impact of sports performance upon wages. After removing the top teams, our estimates showed an effect of back shifted sport performance on wages in the system and difference model, but of no statistical significance.

4.3 Discussion

When summarizing the current dynamic GMM models, we concluded that sports performance, measured in points accumulated over the season positively influences revenues generated per season. We thereby corroborate the findings of Szymanski (1998) as well as those of Pinnuck and Potter (2006) who detected a positive relationship between on-field football performance and off-field financial success. Likewise, Barajas et al. (2005) who examined the relationship between sports performance and revenues in Spanish football, found that sports performance affects revenues of football clubs. A more recent study by Galariotis et al. (2017) also supports such conclusions in French football. The authors claim that better league performance leads to higher revenues because of increased attendance, a greater level of sponsorship, and higher revenues from merchandising, among others. Interestingly, they also found this relationship to exist in the opposite direction: more revenues positively influence sports performance. In our study, we were able to prove the same positive relationship in the reverse direction, i.e., revenues positively impacted the number of points accumulated over the season. Thereby, our results are in agreement with conclusions of Gerrard (2005) who found that the team's revenues are positively related to the league performance of the respective team. Furthermore, in the statistical analyses of H_2, we discovered that sport performance and wage expenditures were positively related. Our current results align with Szymanski (1998) and Rohde and Breuer (2016) that sporting success is driven by team investments.

In order to check for robustness of the current models, we ran the GMM analyses separately on the full data and on the reduced data, where the top six teams were removed. We were regretfully unable to estimate a distinct model with the top six teams only because of computational issues, specifically singularities. When comparing the results of the full and reduced data, the models estimated on full data provided statistically significant hypothesized relationships in the tested as well as the reversed direction, either in the system or the difference GMM model or in both models. The reduced data, however, only yields significant results for H_1 in the system and in the difference GMM model in the tested direction. Significant results were also present in the reversed direction in the difference GMM model. However, no significant results ($\alpha = 0.05$) were found for the reduced data in either direction of H_2.

This might suggest that only financially successful teams that offer high wages to their star players are able to accumulate a larger number of points during the season. It is concluded that the top six teams which generate high revenues also achieve more on-field success. Expectedly, in our study, robustness of the results towards different data subsets remains unproven. Current results suggest that the top teams attract higher attention from fans and sponsors and can thereby generate more revenues through sponsorship and ticketing compared to the mediocre teams. The top teams seem to exert a significant influence upon existence and strength of the explored relationships.

Weak points of this study could be related to the short length of the panel data. For this type of data, longer panels were unavailable, while the number and selection of teams under exploration was determined by the rules of the EPL league. It is further hypothesized that data from teams with home base in the same city, for example Manchester or London, are likely to be extremely correlated thus contributing to the previously mentioned computational problems in some data subsets. Also, we noticed a sizeable sensitivity of the coefficients of respective dynamic GMM models that were related to model specification of the tested relationships.

The relationship that has not been explored in this paper is the one between revenues and wages. Nonetheless, we could assume that there is indeed a noteworthy relationship between revenues and wages: teams with higher revenues are able to spend more on players and coaches, thereby attracting the best talents. Better players consequently shall bring better sports performance as proven in H_2; better sports performance in turn leads to higher revenues as proven in H_1. However, this hypothesized relationship would be potentially relevant and interesting to explore in future research.

5 CONCLUSION

In models verifying H_1 on full data, we detected a positive relationship between sport performance and revenues in both directions. With reduced data, the impact of sports performance upon revenues was statistically established, while the relationship in the reverse direction was only statistically significant in the difference model. Similarly, in dynamic GMM models verifying H_2 on full data, we were able to find a positive relationship between wage expenses and sport performance in both directions. In the reduced data, the impact of wages on sport performance was statistically non-significant and could not be established.

We also conclude that the approach of using panel data is suitable for this sports research since it provides evidence of time effects upon the response variables of interest and explores variable relationships across all teams. The GMM models evidently helped prove the dependencies between the variables with the full data despite existing computational challenges related to unbalanced data. With a more recent dataset and increased sample size, we could be able confirm previous academic research on the relationship between sports performance and revenues as well as player investments and on-field success. Nonetheless, when the top six teams were excluded, results of the model where the independent variable was a financial value, i.e., revenues or wages, were not significant. This observation suggests a sustainable feedback cycle of wages, sport performance and revenues: if sport organizations can afford to pay better players, more on-field success shall follow. And with more success on the pitch, revenues are more likely to increase. Higher revenues shortly mean that more resources become available for players. This virtuous cycle was explored and stated by Baroncelli and Lago (2006). Nonetheless, there are other factors that need to be taken into account, such as strength of the opponent, competitiveness, and economic condition of other leagues.

The relationship between wages and revenues was unexplored in this study, although it would be interesting to include this potential relationship in future research. Further, it would be interesting to verify a relationship between investment and performance, i.e., how much investment needs to be made to achieve the optimal performance and achieve target revenues. More specifically, what an optimal investment looks like so that clubs of smaller financial budget are able to compete not only on the pitch but also in terms of revenue generation. There is certainly plenty to be explored in this field.

6 REFERENCES

Arellano, M. and Bond, S. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58 (2), 277–297. DOI: 10.2307/2297968.

Barajas, Á., Fernández-Jardón, C. M. and Crolley, L. 2005. *Does Sports Performance Influence Revenues and Economic Results in Spanish Football?* Munich Personal RePEc Archive, 3.234. DOI: 10.2139/ssrn.986365.

Baroncelli, A. and Lago, U. 2006. Italian Football. *Journal of Sports Economics*, 7 (1), 13–28. DOI: 10.1177/1527002505282863.

Blundell, R. and Bond, S. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, 87 (1), 115–143. DOI: 10.1016/S0304-4076(98)00009-8.

Blundell, R., Bond, S. and Windmeijer, F. 2000. *Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator*. IFS Working Papers W00/12. Institute for Fiscal Studies, London, UK.

Croissant, Y. and Millo, G. 2018. *Panel Data Econometrics with R*. John Wiley & Sons.

Deloitte. 2020. *Annual Review of Football Finance (editions 2010–2020)* [online]. Available at: `https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/annual-review-of-football-finance.html`. [Accessed 2022, January 15].

Galariotis, E., Germain, C. and Zopounidis, C. 2017. A Combined Methodology for the Concurrent Evaluation of the Business, Financial and Sports Performance of Football Clubs: The Case of France. *Annals of Operations Research*, 266 (1–2), 589–612. DOI: 10.1007/s10479-017-2631-z.

Gerrard, B. 2005. A Resource-Utilization Model of Organizational Efficiency in Professional Sports Team. *Journal of Sport Management*, 19 (2), 143–169. DOI: 10.1123/jsm.19.2.143.

Greene, W. H. 2002. *Econometric Analysis*. 5th ed. New Jersey: Prentice Hall.

Gujarati, D. N., Porter, D. C. and Gunasekar, S. 2012. *Basic Econometrics*. 5th ed. New Delhi: McGraw-Hill.

Guzmán, I. and Morrow, S. 2007. Measuring Efficiency and Productivity in Professional Football Teams: Evidence from the English Premier League. *Central European Journal of Operations Research*, 15 (4), 309–328. DOI: 10.1007/s10100-007-0034-y.

Haas, D. J. 2003. Productive Efficiency of English Football Teams – A Data Envelopment Analysis Approach. *Managerial and Decision Economics*, 24 (5), 403–410. DOI: 10.1002/mde.1105.

Haas, D. J., Kocher, M. G. and Sutter, M. 2004. Measuring Efficiency of German Football Teams by Data Envelopment Analysis. *Central European Journal of Operations Research*, 12 (3), 251–268.

Hadri, K. 2000. Testing for Stationarity in Heterogeneous Panel Data. *The Econometrics Journal*, 3 (2), 148–161. DOI: 10.1111/1368-423X.00043.

Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50 (4), 1029–1054. DOI: 10.2307/1912775.

kicker.de. 2008–2019. *Nuremberg: Olympia-Verlag GmbH* [Online]. Available at: `https://www.kicker.de/premier-league/spieltag/2008-09`. [Accessed 2022, January 15].

Pinnuck, M. and Potter, B. 2006. Impact of On-Field Football Success on the Off-Field Financial Performance of AFL Football Clubs. *Accounting & Finance*, 46 (3), 499–517. DOI: 10.1111/j.1467-629X.2006.00179.x.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing* [online]. R Foundation for Statistical Computing, Vienna, Austria. Available at: `https://www.R-project.org/`.

Ribeiro, A. S. and Lima, F. 2012. Portuguese Football League Efficiency and Players' Wages. *Applied Economics Letters*, 19 (6), 599–602. DOI: 10.1080/13504851.2011.591719.

Rohde, M. and Breuer, C. 2016. Europe's Elite Football: Financial Growth, Sporting Success, Transfer Investment, and Private Majority Investors. *International Journal of Financial Studies*, 4 (2), 12. DOI: 10.3390/ijfs4020012.

Roodman, D. M. 2009a. How to Do Xtabond2: An Introduction to Difference and System GMM in Stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, 9 (1), 86–136. DOI: 10.1177/1536867X0900900106.

Roodman, D. M. 2009b. A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71 (1), 135–158. DOI: 10.1111/j.1468-0084.2008.00542.x.

Sargan, J. D. 1958. The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica*, 26 (3), 393–415. DOI: 10.2307/1907619.

Sloane, P. J. 1971. The Economics of Professional Football: the Football Club as a Utility Maximiser. *Scottish Journal of Political Economy*, 18 (2), 121–146. DOI: 10.1111/j.1467-9485.1971.tb00979.x.

Szymanski, S. 1998. Why is Manchester United So Successful? *Business Strategy Review*, 9 (4), 47–54. DOI: 10.1111/1467-8616.00082.

AUTHOR'S ADDRESS

Marina Schloesser, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: marina.schloesser@icloud.com (corresponding author)

Václav Adamec, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic, e-mail: vadamec@mendelu.cz