sciendo

Handling heteroskedasticity in labour demand functions of athletes

Gábor Rappai
Department of Economics and Econometrics, Faculty of Business and Economics, University of Pécs, Pécs, Hungary
rappai@ktk.pte.hu

Diána Ivett Fűrész
PhD Student, University of Pécs, Pécs, Hungary
furesz.diana@ktk.pte.hu

Abstract:

Based on previous research it can be stated that modelling sport economics related demand curves (e.g. demand for sport events and athletes) is different from other types of modelling. The difference lies in the fact that some parts of the demand curves are nearly horizontal in case of sport goods and nearly vertical in case of athletes, because the price of sport events is inflexible and at the same time, salaries of top athletes are extremely flexible. This study investigates parameter estimation methods appropriate for the relevant demand functions of sport economics. In this cases the generally used ordinary least squares estimator is less robust, so the weighted least squares estimators are able to handle heteroskedasticity. If the distribution of the variables is known, the Newey-West heteroscedasticity corrected estimates give even stronger results. The empirical study analyses footballer transfer fees in top European leagues and identifies a threshold at which the traditional supply-demand functions are not appropriate. According to the results, word class athletes, in a way, can be considered prestige goods for which demand may be irrational.

Keywords: demand curves, football market transfers, robust estimation methods, sports economy.

Introduction

It is a well-known fact in economics that a match between demand and supply leads to a state of equilibrium. This principle is applied throughout the literature concerning job markets in general, where one of the most important tools that are

used in order to achieve or manipulate this state – besides the duration of shifts, the work environment, the overall state of the economy, work experience and talent – is monetary compensation.

Historically, job market analyses have had a primary focus on situations in which supply exceeds demand, resulting in unemployment. Recently, however, the topic of talent scarcity has gained substantial attention. This implies that economic development has been held back by the lack of relevant supply in terms of labour (Davidson, 2009), among other possible factors. From an econometric perspective, this meant the reappearance of disequilibrium models, a concept that seemed to have been extinct for decades. Sports leaders, club owners, coaches and managers, often speak of talent scarcity yet the literature on sports specific job market modelling has not given a satisfactory, in-depth response to this phenomenon to this day. This is especially surprising as it has been almost two decades since Kahn (2000) has labelled this particular job market a textbook example of a laboratory.

This study attempts to describe the unique characteristics of the demand function that is specific to the job market for athletes, and to specify a model that allows its deeper analysis. This involves empirically driven parameter estimations (despite the limited public access to data) and references to some of the unique characteristics of the athletic job market that any industry specific model should incorporate.

Relevant characteristics of the job market within the Sports Industry

According to Primault (2006), approximately 0.7% of the employed population in the European Union earns their income from sports, including those who are routinely referred to as professional athletes. The number of those who are relevant to this analyses is estimated to be around 20,000. This figure calls for a theoretical analysis of the challenges that are triggered by the unique characteristics of the sports industry and the job market within, in particular.

In neoclassical economics, the mechanisms of how equilibrium wages arise and how effective demand appears are explained with the intercept of the demand and supply functions, an approach coined by Alfred Marshall. Figure 1 is a representation of such mechanisms that take place in the job market when there is a shift in aggregate demand. (L is a quantitative representation of labour, and w denotes wages. The numbers in this figure are fictional. Supply (L_S), original and new demand (L_D1, L_D2) can be interpreted as the number of players who signed over from one club to another within a given season).

According to Figure 1, if the demand function (L_D1) is shifted upwards, it only results in a minor increase in wages, where the new demand function (L_D2) is associated with a new equilibrium pay. Such an upward shift can be a result of improved budgetary circumstances, including additional sponsorship, or the introduction of new tax laws subsidizing the industry.

Given the unique characteristics of the job market for athletes, the classical analysis that is based on the intercept of the supply and demand curves may not be appropriate, as it ignores the possibility that the supply curve can become quasi-vertical under realistic conditions (Kesenne, 2014). Figure 2 represents this scenario.
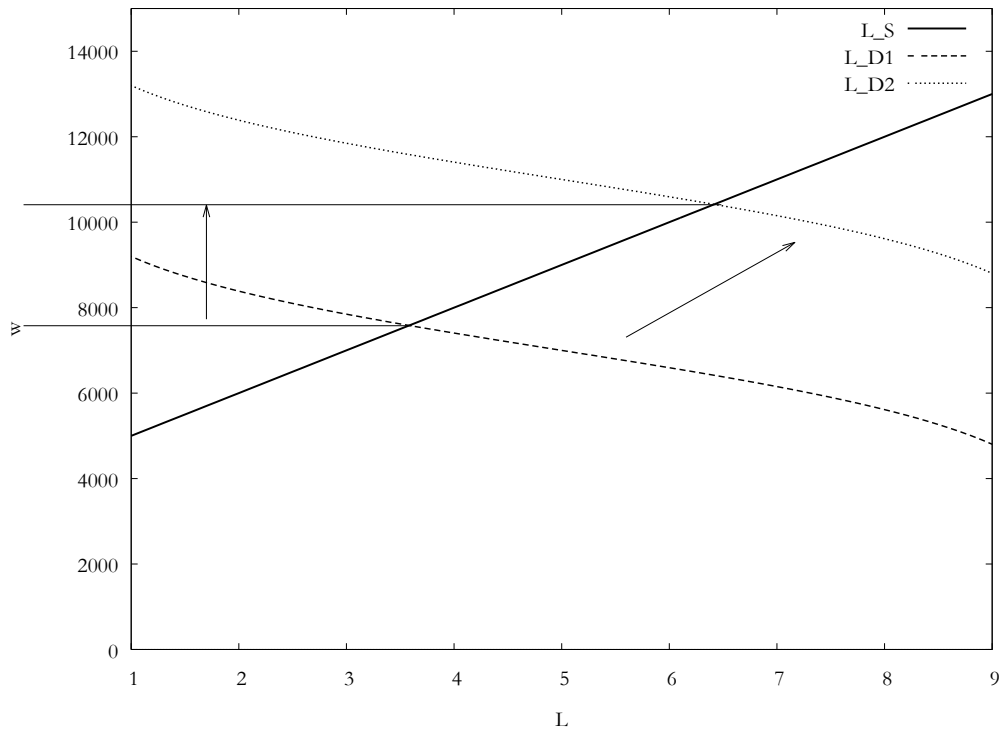
Figure 1 Supply and Demand in a typical Job Market
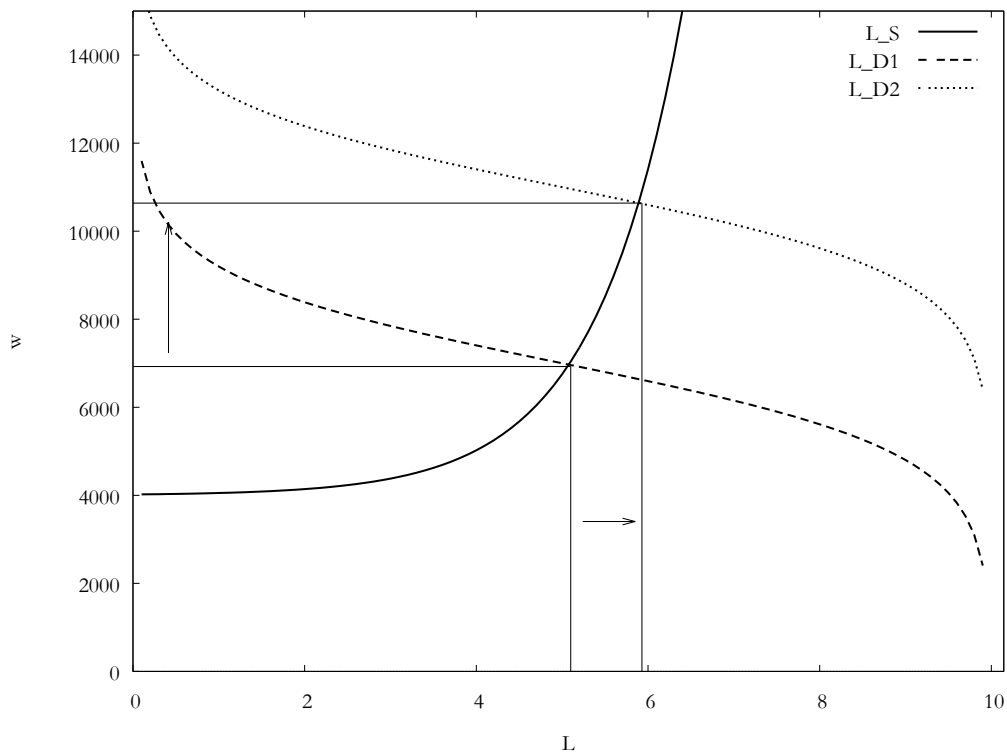Source: Authors' creation.

Figure 2 Demand and Supply, as experienced in the Job Market within the Sports Industry
Source: Authors' creation.

This explains the phenomenon where supply is absorbed by demand and virtually no upward shift of the demand curve enables new employment, regardless of budget. This, from the reverse perspective, explains how athletes respond to bids in the same scenario.

Figure 3 represents how the vertical segment of the labour supply function – the segment of low talent supply elasticity with respect to pay – can cause talent scarcity; in other words, how new jobs cannot be added at virtually any pay rate.

This has consequences even several levels, including the sports clubs, as it can lead to unrealistic salaries, which North American major leagues attempt to fight against with pay caps. Other consequences include a phenomenon where new contracts may not be solely signed with the intent of strengthening one's own club, but also with the purpose of weakening the outlooks for other clubs in terms of their odds of winning. For more in-depth analyses on this phenomenon, see Andreff (2015) or Rosen, Sanderson (2001).
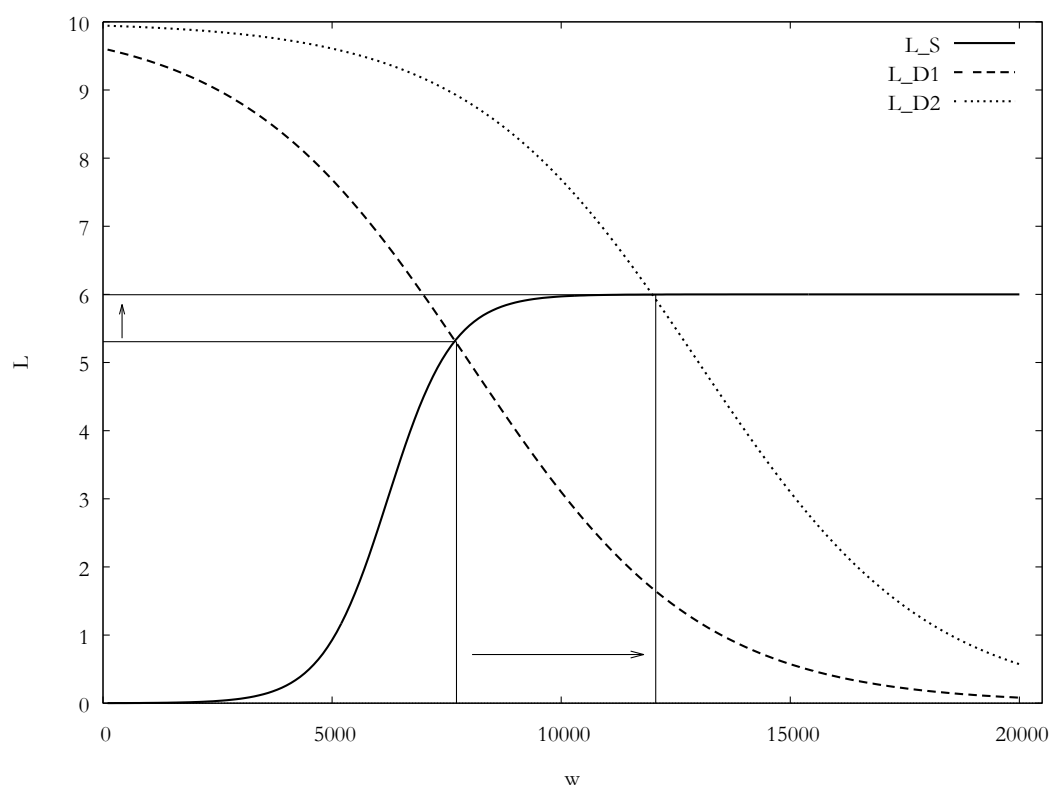

Figure 3 The effect of an upward shift of the demand curve on wages and supply
Source: Authors' creation.

Naturally, it also has unique effects from an econometric perspective, and poses certain challenges in creating well-specified disequilibrium models with the purpose of analysing the labour market in the sports industry. In order to highlight this fact, let us consider the generic structure of disequilibrium models:

$$D_t = \alpha X_t^D + U_t^D$$
$$S_t = \beta X_t^S + U_t^S$$
$$Q_t = \min(D_t, S_t) \tag{1}$$
$$w_t = w_{t-1} + \gamma(D_{t-1} - S_{t-1})$$

Here D_t means demand for athletes at time t, S_t supply of athletes at time t, Q_t number of transactions where athletes sign over to a club at time t, X_t^D, X_t^S explanatory variables of the demand and supply functions, w_t average pay at time t (wage), α, β, γ are model parameters, and u_t^D, u_t^S are residuals.

The most substantial challenge in the process of parameter estimation of disequilibrium models is that demand and supply are not directly observable, as all figures are anticipated. The only observable endogenous variable is the number of transactions, which is always the smaller value of the two. (This statement assumes that the market and the conditions it operates under enable at least one of the two.)

Goldfeld, Quandt (1975) discuss the challenges of model specification and estimation in more detail in Rosen, Quandt (1978).

This study is built on the assumption that any new employment (transaction) is agreed upon near an equilibrium state of supply and demand. It is, in fact, a generous assumption given certain real-life examples; however, it has proven to be a simplification that is acceptable for modelling purposes (Andreff, 2015). From an econometrist's perspective, the real interest lies within the challenge of overcoming the fact that any regression based model that attempts to explain the number of transactions is bound to be heteroskedastic, as talent scarcity increases the chances of unrealistic bids in terms of pay. These figures, therefore, tend to show extreme variability. The empirical section of this study is dedicated to tackling this problem in more detail.

Empirical Demand Curves in the Sports Industry

For the previously discussed reasons, let us assume that demand can be considered a bottleneck in the segment of the job market that is relevant to this analysis, and that the number of new employments (contracts, transactions) between athletes and clubs is near the equilibrium. This, through scarce supply, means that there is a high level of variance in terms of pay, which is what this analysis attempts to model.

Rottenberg (1956) was the first known author to publish about the specifics of the function representing the demand for athletes. Since then, several empirically driven papers have been published with the purpose of refining its specifications (Sloane, 2006), each of which include

- pay (wages), and proxy or surrogate variables that represent
- experience and
- talent (giftedness, skillset),

among the explanatory variables.

Within the last two decades, a number of papers have started to list certain "soft" factors on the right hand side of the equation, such as legal stipulations, pay caps, or racism (Kahn, 2000). Even though these are beyond important variables, their inclusion is not necessary from a modelling or methodological perspective.

The demand function, given the specifics of the industry, can be written as:

$$D_i = f\left(w_i, ex_i, tal_i\right) + \varepsilon_i, \qquad (2)$$

where ex_i represents experience, tal_i is a proxy variable for talent (giftedness), and the residuals are denoted by ε_i. (Each variable is specific to club i.) The next step raises two important questions:

- What surrogate variables can be chosen in order to quantify experience and talent, and
- What type of function can be used in order to realistically map these variables?

Proxy variables are primarily selected based on the types of measurable information that is publically available or accessible. In this study, following the typical choices of such variables, were included

- purchasing price, instead of wages (p, in million Euros)
- the age of the athlete, instead of their direct experience (age, in years), and
- the number of times a given athlete has been chosen to play on the nation team, as a proxy for talent (NTP, as an integer).

Undoubtedly, these may not be the ideal candidates to account for or mediate the actual variables. For example, experience could be measured by the number of years that a given athlete has spent on a champions' league, or by the number of "highly ranked" international games they participated in. Moreover, zero counts in this regard require special attention. (The use of a heckit model would also be a reasonable choice.)

On the other hand, access to such data is either restricted, involve subjective judgement, or are not measured systematically.

Regarding the type, shape and the characteristics of the demand function, the literature is quite heterogeneous, but most authors agree that the relationships between demand and experience and between demand and giftedness are non-linear. It is generally accepted to describe these relationships with logarithmic or power functions where the exponent is less than 1. Moreover, NTP may not fulfil its purpose as a proxy for giftedness, but it is reasonable to introduce a binary variable in order to distinguish between those who had never been on the national team, and those who had.

These considerations lead to a more refined demand function:

$$D_i = \beta_0 + \gamma dum_i^{tal} + \beta_1 p_i + \beta_2 g_1\left(age_i\right) + \beta_3 g_2\left(NTP_i\right) + \varepsilon_i, \qquad (3)$$

where dum_i^{tal} is a binary variable that takes up the value of 1 if a given athlete has ever been hired on to the national team and zero otherwise; $g_1(.)$ and $g_2(.)$ are concave functions to be determined during model specification, and $\beta_0, \beta_1, \beta_2, \beta_3, \gamma$ are the parameters that are to be estimated.

In this modelling experiment, two $g_1(.), g_2(.)$ transformations were applied, and therefore the empirical study ended up testing four different types of demand functions:

a) $D_i = \beta_0 + \gamma dum_i^{tal} + \beta_1 p_i + \beta_2 \sqrt{age_i} + \beta_3 \sqrt{NTP_i} + \varepsilon_i$

b) $D_i = \beta_0 + \gamma dum_i^{tal} + \beta_1 p_i + \beta_2 \log\left(age_i\right) + \beta_3 \sqrt{NTP_i} + \varepsilon_i$

c) $D_i = \beta_0 + \gamma dum_i^{tal} + \beta_1 p_i + \beta_2 \sqrt{age_i} + \beta_3 \log\left(NTP_i\right) + \varepsilon_i$

d) $D_i = \beta_0 + \gamma dum_i^{tal} + \beta_1 p_i + \beta_2 \log\left(age_i\right) + \beta_3 \log\left(NTP_i\right) + \varepsilon_i$

The data source for this experiment was based on the number of transactions that took place within the TOP50 European football clubs within the last ten seasons. Data are from Transfermarkt (2018). For the list of these teams, see the Table A1 in Appendix. That would be interesting to identify different demand functions in case of different positions (e.g. goalkeeper, defender, striker), but the sample size is not sufficiently large enough to do so.

The empirical results are to be interpreted as aggregate (in this case, average) values: accordingly, the dependent variable is the sum of all players who signed over within the last ten seasons; wages are replaced with average purchasing price, and the binary variable represents the number of athletes who have been on their own country's national team. Similarly, NTP and age are the corresponding average values.

The focus is the extent to which the efficiency of the Ordinary Least Squares (OLS) method decreases as the demand function becomes quasi-vertical and causes the residuals to show heteroskedasticity. This latter phenomenon was handled with two well-known methods, the Weighted Least Squares (WLS) method (see Greene, 2008) and the Newey-West estimator (see Newey, West, 1987). Table 1 gives a summary of the estimating functions applied in this analysis.

Table 1 Summary of estimating functions

Methods of Estimation	Estimators
Ordinary Least Squares (OLS)	$\hat{\boldsymbol{\beta}} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T y}$
Weighted Least Squares (WLS) (Therefore, $0 \le dum_i \le D_i$)	$\hat{\boldsymbol{\beta}} = \left(\mathbf{X^T \Omega^{-1} X}\right)^{-1} \mathbf{X^T \Omega^{-1} y}$ $\mathbf{\Omega^{-1}} = \left\langle \dfrac{\sigma^2}{\sigma_1^2}, \dfrac{\sigma^2}{\sigma_2^2}, \ldots, \dfrac{\sigma^2}{\sigma_n^2} \right\rangle$
Newey-West Heteroskedasticity Corrected (NWHC)	$\hat{\boldsymbol{\beta}} = \left(\mathbf{X^T \Omega^{-1} X}\right)^{-1} \mathbf{X^T \Omega^{-1} y}$ $\mathbf{\Omega} = \left\langle \hat{\boldsymbol{\varepsilon}}^T \boldsymbol{\varepsilon} \right\rangle$

Source: Authors' creation.

Using the sample, all parameters were estimated with each of the three methods, and each estimate has undergone the White test (White, 1980), as well. According to this test, the level of heteroskedasticity is considered excessive, if the p value suggests that it is reasonable to reject the null hypothesis or if there is substantial difference between the estimates delivered by the individual methods. Table 2 gives an overview of these estimates and the corresponding tests.

As expected, the White test indicates heteroskedasticity for all OLS estimates. The three methods of estimation, however, do not imply noticeable differences between the models in terms of their goodness of fit or the corresponding F statistics and their p-values. Even though there is little difference between the standard deviations of the parameters, there are substantial gaps between the point estimates provided by the WLS and NWHC methods. It is a noticeable and important result that there are few instances where a significant association could be found between pay and demand, which is likely to be caused by heteroskedasticity. In the rare instances where the estimated $\hat{\beta}_1$ parameter is significantly different from zero, its positivity contradicts the expectation that demand elasticity with respect to price would be negative.

Overall, the best fitting model was d) where the estimates were a result of NWHC, yet the explanatory power is only 8%, and certain parameters were only significant at p=0.05. Based on this model, the top clubs are expected to hire 14 new players within a five-year period, and prefer younger athletes, but are not dedicated to only hiring athletes with experience on a national team.

Table 2 Models, methods and estimates

Model	Method of estimation	Parameter/indicator							
		$\hat{\beta}_0$	$\hat{\gamma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	R^2	F	W
a)	OLS	12.74**	1.07**	4.27	-2.02**	0.81**	0.065	3.37	54.2**
	WLS	43.97**	5.83	1.45	-7.14**	2.45**	0.057	2.99	
	NWHC	0.90**	7.65*	4.60	-1.22**	0.75**	0.060	3.06	
b)	OLS	17.39**	1.08**	4.31	-10.58**	0.80**	0.066	3.38	53.8**
	WLS	6.09**	6.26	1.47	-37.65**	2.45**	0.057	2.92	
	NWHC	11.69**	7.84**	4.58	-6.27**	0.75**	0.060	3.06	
c)	OLS	15.92**	-3.13**	4.84	-1.92**	3.89**	0.076	3.97	53.1**
	WLS	53.96**	-14.57**	1.56	-6.57**	12.28**	0.062	3.20	
	NWHC	11.46**	-1.97**	5.87**	-1.13**	2.91**	0.080	4.19*	
d)	OLS	20.27**	-3.10**	4.87	-10.00**	3.67**	0.076	3.98	52.8**
	WLS	69.53**	-14.45**	1.58	-35.05**	12.25**	0.063	3.21	
	NWHC	13.94**	-1.94**	5.87*	-5.82**	2.91**	0.080	4.20*	

Note: Parameters and test functions marked with ** and * represent that the null hypothesis (by the parameters: the null hypothesis claims that the coefficient is zero, F-test: the null hypothesis claims that the model is not valid, White-test: the null hypothesis claims that the model is homoscedastic) is to be rejected at p=0.01 and p=0.05, respectively.
Source: Authors' creation.

Conclusions

The labour market within the sports industry shows different mechanisms and logic compared to the traditional job markets in the manufacturing or service industry, including how the state of equilibrium is reached. Because of talent scarcity, there may be disproportionate demand for athletes who specialize in popular team sports, which can lead to extreme figures on their paychecks. This analysis had a focus on the challenges of modelling this segment, particularly the implications of the often atypical (horizontal and vertical) shapes of the demand function. It seems that demand models specific to the sport industry need to be capable of handling heteroskedasticity, and that parameter estimates should only be provided after variance reducing transformations. It may be the case that the Newey-West correction is an effective method in this respect. According to the results, word class athletes, in a way, can be considered prestige goods for which demand may be (seemingly) irrational, which may become more explicable if certain causes, besides the traditional logic of economics and market mechanisms, are involved.

References

1. Andreff, W. (2015). *Disequilibrium Sports Economics: Competitive Imbalance and Budget Constraints.* Edward Elgar Publ., Cheltenham.

2. Davidson, E. J. (2009). *Contemporary Models of Giftedness.* In *International Handbook on Giftedness*, Shavinina, L. V. (Ed.) Springer Verlag, Munich, pp. 81-98.
3. Goldfeld, S. M., Quandt, R. E. (1975). Estimation in a disequilibrium model and the value of information. *Journal of Econometrics*, Vol. 3, No. 4, pp. 325-348.
4. Greene, W. H. (2008). *Econometric analysis.* Pearson/Prentice Hall, Upper Saddle River.
5. Kahn, L. M. (2000). The Sports Business as a Labour Market Laboratory. *Journal of Economic Perspectives*, Vol. 14, No. 3, pp. 75-94.
6. Kesenne, S. (2014). *The Economic Theory of Professional Team Sports: An Analytical Treatment.* Edward Elgar Publ., Cheltenham.
7. Newey, W. K., West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, Vol. 55, No. 3, pp. 703-708.
8. Primault, D. (2006). *Employment in sport.* In *Handbook on the Economics of Sport*, Andreff, W., Szymanski, S. (Eds.) Edward Elgar Publ., Cheltenham, pp. 153-167.
9. Rosen, H. S., Quandt, R. E. (1978). Estimation of a Disequilibrium Aggregate Labor Market. *The Review of Economics and Statistics*, Vol. 60, No. 4, pp. 371-379.
10. Rosen, H. S., Sanderson, A. (2001). Labour Markets in Professional Sports. *The Economic Journal*, Vol. 111, No. 469, pp. 47-68.
11. Rottenberg, S. (1956). The Baseball Players' Labor Market. *Journal of Political Economy*, Vol. 64, No. 3, pp. 242-258.
12. Sloane, P. J. (2006). *Rottenberg and the Economics of Sport after 50 Years: An Evaluation.* IZA Discussion Paper No. 2175.
13. Transfermarkt (2018). Transfers & Rumours. Available at http://transfermarkt.com [05 December 2017].
14. White, H. (1980). A Heteroskedasticity – Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, Vol. 48, No. 4, pp. 817-838.

About the authors

Gabor Rappai is the Head of Department of Economics and Econometrics at the Faculty of Business and Economics of the University of Pécs. He teaches Statistics, Multivariate Statistics, Time-series Analysis and Econometrics. Besides his work at the University of Pécs, he is a member of the Economy-Modelling Association and a member of the editing committee of the Hungarian Statistical Review. During his academic career, he authored 86 articles in Hungarian and 15 articles in foreign languages, and participated in the preparation of nine textbooks. He earned his Master's Degree in Economics at the Faculty of Business and Economics of the University of Pécs in 1987, and he received his doctorate in 1989. In 1990, he spent 10 months at the Statistical and Econometrics Institution of the University of Hamburg as a researcher with a DAAD scholarship. In 1993, he held presentations at the Arhuus Business School about financial time-series. He received his Habil from the Faculty of Business and Economics of the University of Pécs in 2003 in the field of econometric modelling. Since 2017, he has been the Doctor of the Hungarian Academy of Science (DSc). His research areas are Econometrics of Causality, Modelling Financial Time Series and Sport Economy. Author can be contacted at: *rappai@ktk.pte hu*.

Diána Ivett Fűrész is a PhD student at the Doctoral School of Health Sciences of the University of Pécs. She has just finished her professional basketball career (80 games in the Hungarian National Team). She has previously graduated in the field of economics. Her scientific research interest is focused on the sports economics supported by her supervisor Professor Gábor Rappai. The main object of her PhD programme is to find the specialities of value creation in sport and the difference between North American and European professional sports. Since 2017 she has been a junior researcher in the Project of Human Resources Management Dent Tender 3.6.2 and she has been teaching statistics at the Faculty of Business and Economics of the University of Pécs. She can be contacted at: *furesz.diana@ktk.pte.hu*.

Appendix

Table A1 TOP50 European football clubs

Juventus	Borussia Dortmund	SSC Napoli	Espanyol Barcelona
AS Roma	FC Chelsea	FC Arsenal	FC Getafe
Tottenham Hotsp.	Olympique Lyon	Spartak Moscow	FC Southampton
Inter Milan	Olymp. Marseille	Eintracht Frankfurt	Galatasaray Istanbul
Udinese	Werder Bremen	FC Paris Saint Germain	FC Nantes
VFB Stuttgart	AFC Sunderland	Newcastle United	SS Lazio
FC Liverpool	Bayer Leverkusen	West Ham United	FC Bordeaux
AC Milan	Real Madrid	Torino FC	FC Everton
FC Sevilla	FC Schalke 04	FC Barcelona	Leicester City
ACF Fiorentina	FC Valencia	FC Bayern Munchen	Zenit St. Petersburg
AS Monaco	FC Villarreal	OGC Nizza	CSKA Moscow
Manchester Unit.	Manchester City	B. Mönchengladbach	Fenerbahce Istanbul
Atletico Madrid			FC Girona

Source: Transfermarkt (2018).