Evaluating professional tennis players' career performance: A Data Envelopment Analysis approach

George Halkos and Nickolaos Tzeremes

University of Thessaly, Department of Economics

September 2012

Evaluating professional tennis players' career performance: A Data Envelopment Analysis approach

By

George E. Halkos and Nickolaos G. Tzeremes

University of Thessaly, Department of Economics,

Korai 43, 38333, Volos, Greece

Abstract
This paper by applying a sporting production function evaluates 229 professional tennis players' career performance. By applying Data Envelopment Analysis (DEA) the paper produces a unified measure of nine performance indicators into a single career performance index. In addition bootstrap techniques have been applied for bias correction and the construction of confidence intervals of the efficiency estimates. The results reveal a highly competitive environment among the tennis players with thirty nine tennis players appearing to be efficient.

Keywords: Professional tennis players; Data Envelopment Analysis; Sport production function; Bootstrapping.

JEL classification: C14**;** C69**;** L83.

1. Introduction

The economic theory behind sporting activity is based on the work of Rottenberg (1956). However, Scully (1974) was the first to apply a production function in order to provide empirical evidence for the performance of baseball players. Since then several scholars have used frontier production function in order to measure teams' performance and which has been described on the works of Zak, Huang and Siegfried (1979), Porter and Scully (1982) and Fizel and D'Itri (1996, 1997).

Similarly, Dawson, Dobson and Gerrard (2000a) applied stochastic frontier approach (SFA), measuring managers' efficiency for a panel of managers in English Football (soccer) Premier league. The study by Haas (2003a) was one of the first studies which applied data envelopment analysis (DEA) in order to measure team efficiency of twenty English Premier League clubs. More recently, Barros and Leach (2006a, 2006b, 2007) applied a stochastic Cobb-Douglas production frontier and DEA in order to measure the performance of football clubs in the English F.A. Premier League.

Even though several studies have used DEA methodology to evaluate mostly football teams' efficiency levels, there are not any studies proposing a similar production function approach for evaluating professional tennis players' efficiency levels. Recently a study by Rámon, Ruiz and Sirvent (2012) proposed a DEA model with no inputs and a common set of weights in order to evaluate professional tennis players' efficiency levels. In contrast to the pre-mentioned study this study uses a sports production function approach in a DEA framework in order to evaluate 229 professional tennis players' career efficiency levels. In that respect our paper contributes to the existing literature by providing a DEA based multi-criteria indicator of tennis players' performance evaluation.

The structure of our paper is the following. Sections two and three present respectively a brief literature review and the data and the methodology adopted in the study respectively. Section four discusses the derived empirical results, whereas the last section concludes the paper.

2. Literature review

The production function in professional team sports is based on full technical efficiency in a sense that output can be maximized given a level of inputs or the inputs can be minimized given a set of outputs[1]. However, according to Dawson, Dobson and Gerrard (2000b) professional teams are not behaving as profit maximisers due to the existence of principal-agent relationships and therefore, fail to minimize costs and thus to be technical efficient.

According to Rottenberg (1956) the "product" in sporting production function is the admission revenue generated by the sporting contest but, according to Dawson, Dobson and Gerrard (2000b) the modeling idea of treating revenues as an output have not been supported by researchers in the field of sport economics. However, Scully's (1974) specification of treating as input individual players performance and team performance as output became an acceptable modeling strategy of the sporting production function.

According to Dawson, Dobson and Gerrard (2000a, 2000b) there are two important limitations for the studies following this approach. First, the use of OLS regression analysis represents the average efficiency rather than the absolute creating

[1] For an extensive analysis and literature review of sports production function see Dawson, Dobson and Gerrard (2000b) and Lee and Berri (2008).

several measurement problems for policy evaluation[2]. Secondly, Scully's production function excludes the impact of coaches which according to several studies play a major role of teams' performance (Zech, 1981; Carmichael & Thomas, 1995).

However, to our opinion still there are other aspects affecting directly teams' performance and can not be captured due to lack of information. These may well be the history of the team, the spirit or other capital and labor related factors such as the physiatrists, doctors, training staff, training centers and their explicit personnel, youth academies and their personnel, etc.

Several studies have applied efficiency analysis on sport teams' performances[3]. Though the economic framework of professional sporting activity is based on the works of Rottenberg (1956), Neale (1964), Jones (1969) and Sloane (1969, 1971, 1976). In addition the first empirical evidence in an average production function framework was found in the work of Scully (1974) who investigated the performance of baseball players. By using the percentage of matches won in order to model teams' output and management, capital and team spirit as inputs, Scully's empirical work was the first to apply a production function in order to provide empirical evidence. The sporting production process has been modeled by several others in a similar way (among others Zech, 1981; Atkinson, Stanley & Tschirart 1988; Schofield, 1988).

The application of frontier production function in order to measure teams' performance has been dated back on the works of Zak, Huang and Siegfried (1979), Porter and Scully (1982) and Fizel and D'Itri (1996, 1997). Additionally, over the last

[2] Dawson, Dobson and Gerrard (2000b) suggest that stochastic frontier approach (SFA) is more suitable for measuring teams' performance compared to OLS approach. However as noted by Lee and Berri (2008) the time-varying stochastic frontier models with the identical temporal pattern assumption cannot be used in the analysis of team efficiency in sports.

[3] For a literature review on the subject matter see Barros and Garcia-del-Barrio (2008).

two decades several scholars have been applying parametric and nonparametric frontier analysis to establish football teams' performance and their determinants. Dawson, Dobson and Gerrard (2000a), applying stochastic frontier approach (SFA), measure managers' efficiency for a panel of managers in English soccer's Premier league using as output the percentage of matches won and as inputs several player quality variables, for the time period of 1992 to 1998.

Haas (2003b) applied a data envelopment analysis (DEA) measuring team efficiency of the USA Major League Soccer (MLS). In a DEA setting and for the year 2000, Haas used head coaches' and players' wages as inputs and revenues, points awarded and number of spectators as outputs. In addition Haas (2003a) in a similar DEA setting performed an efficiency analysis for twenty English Premier League clubs for the year 2000-2001. Furthermore, Barros and Leach (2006a, 2006b, 2007) applying a stochastic Cobb-Douglas production frontier and DEA measured the performance of football clubs in the English F.A. Premier League for the time periods 1989-1990 to 2002-2003. They applied a combination of sport and financial data in order to measure football clubs' efficiency levels.

Frick and Simmons (2008) by applying SFA on data for German premier soccer league (Bundesliga) showed that managerial compensation impact positively on team success. In addition using the recent developments of efficiency measurement Barros, Assaf and Sá-Earp (2010) by applying Simar and Wilson's (2007) DEA bootstrap procedure analyzed the performance of the Brazilian first league football clubs. Similarly, Barros and Garcia-del-Barrio (2011) measured the efficiency of the Spanish football clubs for the seasons 1996–1997 and 2003–2004 by applying the two-stage procedure (Simar & Wilson, 2007).

In terms of professional tennis related studies, several of them have concentrated on predicting tennis matches' outcomes. For instance, del Corral and Prieto-Rodríguez (2010) by estimating probit models and applying bootstrapping techniques using Grand Slam tennis match data from 2005 to 2008 test whether the differences in rankings between individual players are good predictors for Grand Slam tennis outcomes. Several other predicting models for tennis matches outcomes have been also presented over the years (Boulier & Stekler, 1999; Clarke & Dyte, 2000; Klaassen & Magnus, 2003; Scheibehenne & Broder, 2007; McHale & Morton, 2011).

Moreover, in a different study Coate and Robbins (2001) haven't found any evidence that top-ranked male tennis professionals are more dedicated to their careers compared to the top-ranked female professionals. In addition Wozniak (2012) by using data from the International Tennis Federation (ITF) provides evidence that gender differences are smaller in a very competitive setting and that the effects of previous performance are dependent on gender and the time the previous competition took place. Rohm, Chatterjee and Habibullah (2004) proposed an approach for measuring competitive dynamics for major tennis championships providing evidences that competitiveness at Wimbledon has been extremely high.

Nevertheless, the only study applying DEA methodology in order to evaluate professional tennis players' efficiency levels is the one conducted by Rámon, Ruiz and Sirvent (2012). By using data from Association of Tennis Professional (ATP) they evaluated the efficiency level of 53 professional tennis players who played more than 40 matches. Rámon, Ruiz and Sirvent (2012) proposed a DEA model with no input specifications and nine performance outputs using the assumption of constant returns to scale (CRS) and a common set of weights. As a result their derived final player ranking was similar to the one proposed from the ATP.

Finally, in contrast to the pre-mentioned study, our study uses a sports production function setting in a DEA context evaluating career performance of 229 professional tennis players. Moreover, different economic assumptions have been used in our setting and bootstrapped techniques have also been applied in order for our results to be corrected from sample bias.

3. Data and Methodology

3.1 Description of the variables

We are using career data from the Association of Tennis Professional (ATP) for 229 world tennis players (see appendix A1) for the time period between January 1991 and July 2012[4]. The ATP data concern 10 performance factors recorded for three types of tennis courts (clay, grass and hard courts) such as: career matches (input), career break points saved (output), career aces (output), career 1st serve points won (output), career 2nd serve points won (output), career service games won (output), career 1st serve return points won (output), career 2nd serve return points won (output), career break points converted-points won (output) and career return games won (output). We are using these factors in a sports' production function framework having one input (career matches) and nine different outputs (performance determinants).

According to Rámon, Ruiz and Sirvent (2012, p.4885) the factors provided from the ATP rankings are based on players' performances on world tournaments during different seasons by providing to the players a different amount of points based on the rounds they reach in such tournaments. Therefore, ATP rankings are determined according to the total points a tennis player gets in a specific season. As

[4] The data are available from the official site of the Association of Tennis Professional (ATP) at: http://www.atpworldtour.com/

such they reflect players' competitive position but in contrast to our study do not reflect players' efficiency performance of their career games[5].

As can be observed in our setting of tennis players' production function we are using one input (career matches) and nine performance indicators. Figure 1 provides us with kernel conditional density estimate using local polynomial estimation following Hyndman, Bashtannyk and Grunwald (1996) and Hyndman and Yao (2002)[6]. The conditional density plots are computed for tennis players' nine different output performance measures[7] and their career matches. Unlike standard descriptive tables, the density plots can provide us with complete picture of the underlying processes generating the selected outputs. All subfigures reveal a positive relationship between tennis players' career matches and the nine selected output performance indicators. Indicating that as the input performance increase (career matches) the output performance indicators are also increasing.

Besides, Figure 1 presents large dispersions among the nine different outputs used. For instance, a large dispersion at high number of career matches in the relationship between Career break points saved and career matches is reported (subfigure 1a). In addition for the career aces and career matches relationship the dispersion is more pronounced with evidence of bi-modality (subfigure 2b). Dispersions among the input and the outputs used can also be observed in most of the cases.

---

[5] Scheibehenne and Broder (2007) provided evidence that the official ATP rankings do not provide useful information predicting the outcomes of tennis matches.

[6] The bandwidths have been computed using normal reference rules described in Bashtannyk and Hyndman (2001) and Hyndman and Yao (2002).

[7] These are: career break points saved-subfigure 1a, career aces-subfigure 1b, career 1st serve points won-subfigure 1c, career 2nd serve points won-subfigure 1d, career service games won-subfigure 1e, career 1st serve return points won-subfigure 1f, career 2nd serve return points won-subfigure 1g, career break points converted-points won-subfigure 1h and career return games won-subfigure 1i.

Moreover, in accordance with figure 1, table 1 provides the descriptive statistics of the input and the outputs used. Finally, it can be realized when looking at the high standard deviation values there are a lot of dissimilarities among the output performance indicators of the 229 professional tennis players.

Figure 1: Conditional density plots of tennis players' career matches and output performance determinants

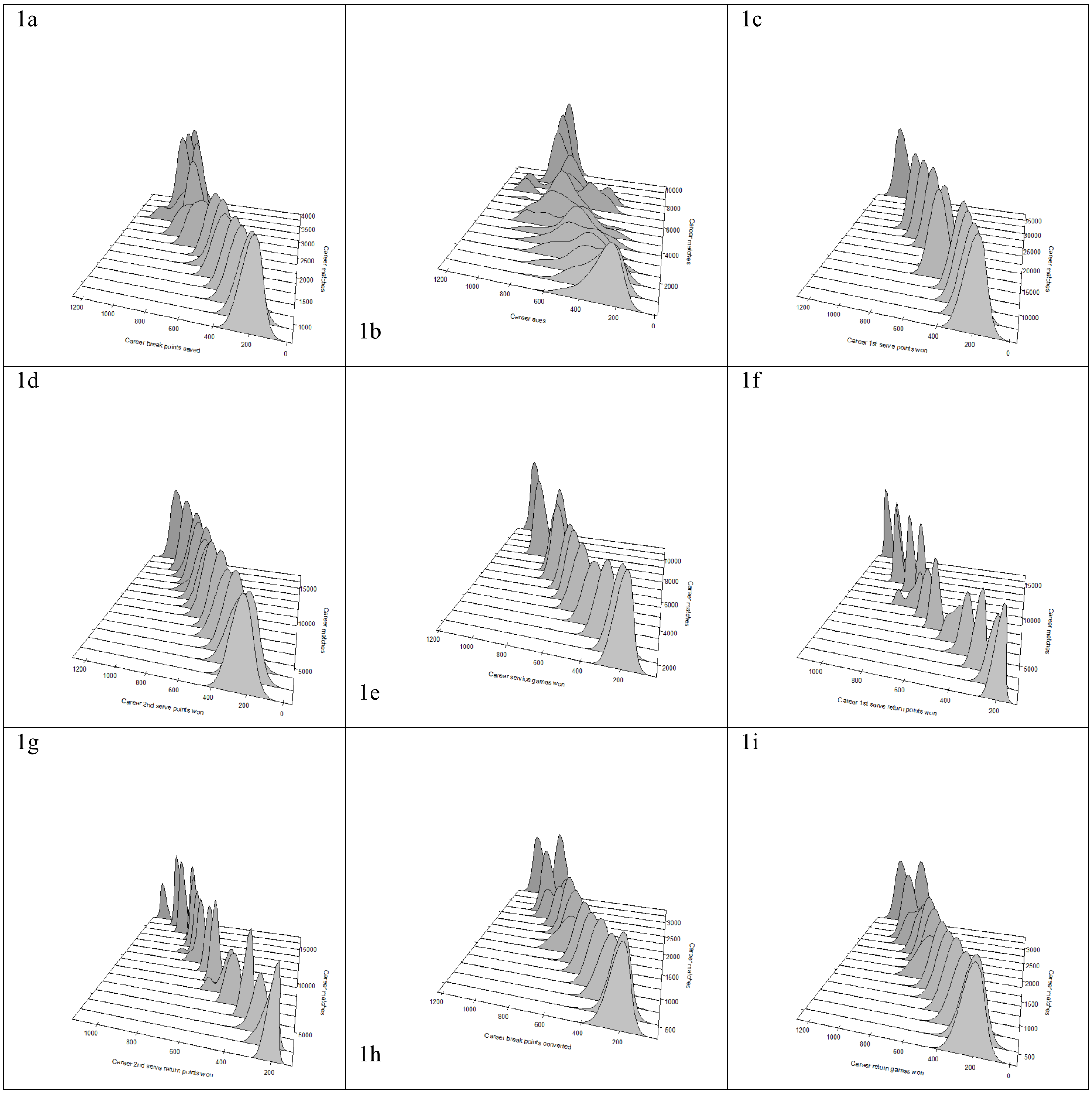


Table 1: Descriptive statistics of the input and outputs used (one input and nine outputs)

	Career matches -Input	Career break points saved-Output	Career aces-Output
Mean	433.454	1678.271	2363.904
Std	181.090	666.213	1689.066
Min	200.000	668.000	248.000
Max	1045.000	4008.000	10183.000
	Career service games won-Output	**Career 1st serve return points won-Output**	**Career 2nd serve return points won-Output**
Mean	3953.148	5848.258	6789.956
Std	1822.707	2625.554	3119.360
Min	1380.000	1932.000	2352.000
Max	11272.000	16344.000	16743.000
	Career 1st serve points won-Output	**Career 2nd serve points won-Output**	**Career break points converted-Output**
Mean	13893.476	6675.681	1228.913
Std	6128.319	3033.200	621.132
Min	5113.000	2050.000	306.000
Max	36871.000	16931.000	3390.000
	Career return games won-Output		
Mean	1228.913		
Std	621.132		
Min	306.000		
Max	3390.000		

3.2 The economic model

Having a list of inputs p and outputs q in an economic analysis framework the performance evaluation of tennis players in a sense of a productivity unit can be defined as a set of points Ψ , which in turn defines the tennis players' production set in the Euclidean space R_{+}^{p+q} as[8]:

$$\Psi = \{(x, y) x \in | R_{+}^{p}, y \in R_{+}^{q}, (x, y) \text{ is feasible}\} \quad (1),$$

the input vector is indicated by x and the output vector by y . Then the output correspondence set (for all $x \in \Psi$) can be defined as:

$$P(x) = \{y \in R_{+}^{q} | (x, y) \in \Psi\} \quad (2).$$

According to Farrell (1957) the efficient boundaries of Ψ can be defined in the output space in radial terms as:

[8] In this section we follow the notation by Daraio and Simar (2007).

$$\partial P(x) = \{y | y \in P(x), \lambda(y) \notin P(x), \forall \lambda > 1\} \quad (3).$$

In addition following Shephard (1970) we assume the economic axioms of no free lunch, free disposability, bounded, closeness and convexity. Then we can define the efficiency level of a tennis player operating at the level of (x_0, y_0) by calculating the distance of these points to the frontiers. Based on the work of Debreu (1951), Farrell (1957) suggested a radial distance in order for the distance from a point to the corresponding frontier to be calculated. In our case we define the frontier in the output direction where the efficient subset is characterized by $\partial P(x)$. Then Farrell's output measure of efficiency can be defined as:

$$\lambda(x_0, y_0) = \sup\{\lambda | \lambda y_0 \in P(x_0)\} = \sup\{\lambda | (x_0, \lambda y_0) \in \Psi\} \quad (4).$$

According to Daraio and Simar (2007) $\lambda(x_0, y_0) \geq 1$ is the proportionate increase of output which a tennis player should achieved in order to be determined as efficient[9].

Finally, we are using the inverse of these radial distances known as Shephard's output distance functions (Shephard, 1970) and can be defined as:

$$\delta^{out}(x, y) = \inf\{\lambda > 0 | (x, \lambda^{-1} y) \in \Psi\} \equiv (\lambda(x, y))^{-1} \quad (5),$$

for all $(x, y) \in \Psi$, $\delta^{out}(x, y) \leq 1$, then for a $\delta^{out}(x, y) = 1$, then (x, y) belongs to the frontier Ψ.

3.3 The Data Envelopment Analysis (DEA) estimator

Farrell (1957) was the first to indicate that linear programming can be used in order to find the frontier and estimate efficiency score. However the linear programming estimators were first operationalized by Charnes, Cooper and Rhodes (1978) presenting a linear program (known as the CCR model) for calculating

[9] This implies that $(x_0, \lambda(x_0, y_0) y_0)$ is on the frontier.

efficiency scores under the assumption of constant returns to scale (CRS)[10]. Later, Banker, Charnes and Cooper (1984) introduced a DEA estimator (known as the BCC model) allowing for variable returns to scale (VRS). The measurement of efficiency for a given tennis player (x, y) relative to the boundary of the convex hull of $\mathrm{X} = \{(X_i, Y_i), i = 1,...,n\}$ can be defined as:

$$\hat{\Psi}_{DEA} = \left\{ (x, y) \in R_{+}^{p+q} \middle| y \le \sum_{i=1}^{n} \xi_i Y_i; x \ge \sum_{i=1}^{n} \xi_i X_i, for(\xi_1,...,\xi_n) \right. \\ \left. s.t. \sum_{i=1}^{n} \xi_i = 1; \xi_i \ge 0, i = 1,...,n \right\} \quad (6).$$

The $\hat{\Psi}_{DEA}$ allows variable returns to scale. However, if we want to allow constant returns to scale the restriction $\sum_{i=1}^{n} \xi_i = 1$ needs to be dropped from equation (6). Then the DEA estimator of the output efficiency score for a given tennis player (x_0, y_0) is obtained by solving the following linear program:

$$\hat{\lambda}_{DEA}(x_0, y_0) = \sup\left\{ \lambda \middle| (x_0, \lambda y_0) \in \hat{\Psi}_{DEA} \right\} \quad (7)$$

$$\hat{\lambda}_{DEA}(x_0, y_0) = \max\left\{ \lambda \middle| \lambda y_0 \le \sum_{i=1}^{n} \xi_i Y_i; x_0 \ge \sum_{i=1}^{n} \xi_i X_i, \lambda > 0; \right. \\ \left. s.t. \sum_{i=1}^{n} \xi_i = 1; \xi_i \ge 0, i = 1,...,n \right\} \quad (8)$$

In addition by applying the methodology introduced by Simar and Wilson (1998, 2000a, 2000b) we perform the bootstrap procedure for DEA estimators in order to obtain biased corrected results (see appendix A2 for details). More analytically the biased corrected estimations can be obtained from:

[10] For interesting remarks raised regarding the history and originality of DEA models see the work of Førsund and Sarafoglou (2002) and Førsund, Kittelsen and Krivonozhko (2009).

$$\hat{\hat{\lambda}}_{DEA}(x_0, y_0) = \hat{\lambda}_{DEA}(x_0, y_0) - \hat{bias}_B\left(\hat{\lambda}_{DEA}(x_0, y_0)\right)$$
$$= 2\hat{\lambda}_{DEA}(x_0, y_0) - B^{-1}\sum_{b=1}^{B=2000}\hat{\lambda}^{*}_{DEA,b}(x_0, y_0) \qquad (9)$$

Then by expressing the output oriented efficiency in terms of the Shephard (1970) output distance function as $\hat{\delta}^{out}_{DEA}(x_0, y_0) = (\lambda(x_0, y_0))^{-1}$ we can construct bootstrap confidence intervals for $\hat{\delta}^{out}_{DEA}(x_0, y_0)$ as:

$$\hat{\delta}^{out}_{DEA}(x_0, y_0) - \hat{\alpha}_{1-\alpha/2}, \hat{\delta}^{out}_{DEA}(x_0, y_0) - \hat{\alpha}_{\alpha/2} \qquad (10).$$

Finally we follow the bootstrap test developed by Simar and Wilson (2002) in order to test whether the CRS or VRS formulation is appropriate to our analysis. The null hypothesis of the test can be developed as $H_0 : \Psi^{\lambda}$ is globally CRS against $H_1 : \Psi^{\lambda}$ is globally VRS. Then the test statistic mean of the ratios of the efficiency scores is then provided by:

$$T(X_n) = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\lambda}_{CRS,n}(X_i, Y_i)}{\hat{\lambda}_{VRS,n}(X_i, Y_i)} \qquad (11).$$

At the same time the p-value of the null-hypothesis can be obtained as:

$$p - value = prob(T(X_n) \leq T_{obs} | H_0 \quad is \quad true) \qquad (12)$$

where T_{obs} is the value of T computed on the original observed sample X_n.The p-value can be approximated by the proportion of bootstrap values of T^{*b} less the original observed value of T_{obs} such as:

$$p - value \approx \sum_{b=1}^{B}\frac{\Im\left(T^{*b} \leq T_{obs}\right)}{B} \qquad (13).$$

4. Empirical findings

According to O' Donnell (2012) the main advantages of DEA is the construction of a surface that envelops the data points taking into account the underline assumptions of the production technology (i.e. convexity) having no assumptions regarding the functional form of the production frontier. On top, DEA can accommodate multiple inputs and outputs with no assumptions of the error terms, no statistical issues (e.g., endogeneity) associated with estimating technologies and it does not require a priori assignments of weights to tennis players' performance dimensions used in order to evaluate their career efficiency levels (Barros, 2003).

Nevertheless, the main weaknesses lie on the fact that DEA can not distinguish inefficiency from noise, it is sensitive to outliers and for small samples the results are biased upwards. Simar and Wislon (1998, 2000a, 2000b) proposed a bootstrap method for inference about the obtained DEA estimators providing in such a way biased free estimators in a multivariate framework.

In our case the DEA methodology is more suitable compared to the econometric approach of efficiency measurement (i.e. stochastic frontier approach-SFA) since we do not know the underlying functional form of tennis players' sport production function and since we combine in our analysis multiple outputs (nine output performance measures) which are difficult to be modelled in a SFA framework.

In addition as has been previously analysed we are using an output orientation in our analysis. Therefore, tennis players' efficiency is based on the ability of the players to expand proportionally their outputs quantities without altering the input quantities used. Coelli, Rao, O'Donnell and Battese (2005) suggest that the choice of orientation is based according to which quantities (inputs or outputs) the decision maker has most control over. As a result, in our case the tennis players have a better control (due to their skills possessed) of their output performance indicators compared to the number of their career matches (input variable)[11].

Table 2 presents the efficiency scores of the professional tennis players assuming constant returns to scale (CRS). The results present that thirty nine players are reported to be efficient (i.e. with efficiency score equal to 1) and a hundred and ninety tennis players are reported to be inefficient (i.e. efficiency scores less than 1). Moreover table 2 presents also the thirty nine players with the lowest performance over the years. In addition to the original estimates, the third column presents the biased corrected estimates (BC_{CRS}), fourth the estimated bias (Bias), the fifth column presents the standard deviation of the bias (Std) and the sixth and seventh columns present the lower (LB) and upper (UB) bounds of the 95% bootstrap intervals following the methodology by Simar and Wilson (1998, 2000a, 2000b).

It is reported that for the efficient tennis players under the assumption of CRS the largest deviation between the original and the biased corrected estimates have been reported for Roger Federer, Patrick McEnroe, Todd Martin, Lleyton Hewitt, Richard Krajicek, Andy Murray, Guillermo Coria, John Isner, Ivo Karlovic, and Wayne Arthurs.

Similarly, table 3 presents the results obtained under the variable returns to scale assumption (VRS). In this case sixty four tennis players are reported to be efficient (i.e. efficiency score equal to 1) and a hundred and sixty five players are reported to be inefficient (i.e. efficiency scores less than 1). Moreover, table 3 presents the last sixty four performers along side with their biased corrected efficiency estimates and 95% lower and upper bootstrapped confidence intervals. In addition the

[11] Nevertheless it must be mentioned that under the assumption of constant returns to scale the output and input oriented measures of technical efficiency are equivalent (Färe and Lovell, 1978)

higher deviations between the original and the biased corrected efficiency estimates under the VRS assumption (BC_{VRS}) are reported for Ivan Lendl, Guillermo Coria, Wayne Arthurs, Fabrice Santoro, Roger Federer, Paul Goldstein, John Isner, Goran Ivanisevic, Ivo Karlovic and Jeremy Chardy.

Table 2: Original and bias corrected efficiency estimates along side with their 95% bootstrap intervals based on the CRS assumption

Players	CRS	BCcrs	Bias	Std	LB	UB	Players	CRS	BCcrs	Bias	Std	LB	UB
Aaron Krickstein	1.000	0.974	0.026	0.000	0.962	0.998	Hernan Gumy	0.898	0.889	0.009	0.000	0.883	0.896
Alberto Berasategui	1.000	0.970	0.030	0.000	0.953	0.998	Igor Kunitsyn	0.898	0.883	0.015	0.000	0.871	0.897
Alex O'Brien	1.000	0.977	0.023	0.000	0.965	0.998	Renzo Furlan	0.898	0.888	0.010	0.000	0.880	0.896
Andre Agassi	1.000	0.962	0.038	0.000	0.942	0.999	Marcelo Rios	0.895	0.878	0.017	0.000	0.863	0.893
Andrew Ilie	1.000	0.964	0.036	0.000	0.941	0.998	Adrian Voinea	0.894	0.880	0.015	0.000	0.871	0.893
Andy Murray	1.000	0.949	0.051	0.001	0.914	0.998	Max Mirnyi	0.894	0.880	0.015	0.000	0.866	0.893
Bernd Karbacher	1.000	0.974	0.026	0.000	0.960	0.998	Benjamin Becker	0.894	0.875	0.019	0.000	0.857	0.892
Boris Becker	1.000	0.974	0.026	0.000	0.952	0.998	Andrea Gaudenzi	0.892	0.878	0.013	0.000	0.869	0.890
Brad Gilbert	1.000	0.981	0.019	0.000	0.971	0.998	Karim Alami	0.891	0.879	0.012	0.000	0.870	0.889
Bryan Shelton	1.000	0.971	0.029	0.000	0.956	0.998	Fernando Meligeni	0.891	0.880	0.011	0.000	0.873	0.889
David Wheaton	1.000	0.965	0.035	0.000	0.949	0.998	Ernests Gulbis	0.891	0.876	0.015	0.000	0.863	0.889
Dmitry Tursunov	1.000	0.983	0.017	0.000	0.974	0.998	Jerome Golmard	0.890	0.879	0.010	0.000	0.872	0.888
Fernando Verdasco	1.000	0.976	0.024	0.000	0.965	0.998	Byron Black	0.889	0.872	0.017	0.000	0.855	0.888
Fernando Vicente	1.000	0.977	0.023	0.000	0.964	0.998	Brett Steven	0.889	0.877	0.011	0.000	0.870	0.887
Gael Monfils	1.000	0.965	0.035	0.000	0.940	0.999	Daniel Vacek	0.889	0.874	0.014	0.000	0.863	0.887
Gianluca Pozzi	1.000	0.972	0.028	0.000	0.951	0.998	Stefan Koubek	0.886	0.876	0.010	0.000	0.869	0.884
Guillermo Coria	1.000	0.944	0.056	0.001	0.902	0.998	Raemon Sluiter	0.884	0.873	0.011	0.000	0.862	0.883
Hendrik Dreekmann	1.000	0.962	0.038	0.000	0.940	0.998	Nicolas Massu	0.883	0.871	0.012	0.000	0.864	0.881
Ivan Lendl	1.000	0.970	0.030	0.000	0.948	0.998	Kenneth Carlsen	0.882	0.868	0.014	0.000	0.857	0.880
Ivo Karlovic	1.000	0.939	0.061	0.003	0.865	0.998	Mariano Puerta	0.877	0.866	0.011	0.000	0.858	0.875
Jan-Michael Gambill	1.000	0.969	0.031	0.000	0.952	0.998	Marc Rosset	0.875	0.859	0.016	0.000	0.846	0.874
John Isner	1.000	0.944	0.056	0.002	0.887	0.998	Janko Tipsarevic	0.869	0.856	0.013	0.000	0.845	0.867
Lleyton Hewitt	1.000	0.959	0.041	0.001	0.929	0.998	Andrei Pavel	0.865	0.852	0.013	0.000	0.842	0.864
Marat Safin	1.000	0.976	0.024	0.000	0.961	0.999	Sebastien Lareau	0.864	0.851	0.013	0.000	0.839	0.862
Mark Philippoussis	1.000	0.967	0.033	0.000	0.944	0.998	Paradorn Srichaphan	0.863	0.851	0.012	0.000	0.843	0.861
Michael Chang	1.000	0.966	0.034	0.000	0.946	0.998	Florian Mayer	0.862	0.853	0.009	0.000	0.849	0.860
Mikael Tillstrom	1.000	0.977	0.023	0.000	0.962	0.998	Kristof Vliegen	0.855	0.847	0.009	0.000	0.841	0.853
Novak Djokovic	1.000	0.975	0.025	0.000	0.957	0.998	Marcelo Filippini	0.851	0.836	0.015	0.000	0.824	0.850
Patrick McEnroe	1.000	0.961	0.039	0.000	0.937	0.998	Filip Dewulf	0.851	0.839	0.012	0.000	0.830	0.849
Paul Goldstein	1.000	0.971	0.029	0.000	0.946	0.999	Christophe Rochus	0.848	0.826	0.022	0.000	0.805	0.846
Pete Sampras	1.000	0.962	0.038	0.000	0.936	0.998	Carl-Uwe Steeb	0.845	0.838	0.007	0.000	0.833	0.844
Rafael Nadal	1.000	0.966	0.034	0.000	0.942	0.998	Shuzo Matsuoka	0.845	0.830	0.015	0.000	0.818	0.843
Richard Fromberg	1.000	0.983	0.017	0.000	0.970	0.999	Christian Ruud	0.841	0.828	0.012	0.000	0.819	0.839
Richard Krajicek	1.000	0.958	0.042	0.001	0.930	0.998	Jakob Hlasek	0.838	0.826	0.012	0.000	0.818	0.836
Roger Federer	1.000	0.961	0.039	0.000	0.940	0.999	Yen-Hsun Lu	0.834	0.822	0.012	0.000	0.813	0.833
Scott Draper	1.000	0.972	0.028	0.000	0.955	0.998	Daniel Nestor	0.821	0.805	0.016	0.000	0.792	0.820
Todd Martin	1.000	0.961	0.039	0.000	0.937	0.998	Hyung-Taik Lee	0.805	0.799	0.007	0.000	0.794	0.804
Tomas Carbonell	1.000	0.969	0.031	0.000	0.954	0.998	Luis Horna	0.798	0.785	0.013	0.000	0.773	0.796
Wayne Arthurs	1.000	0.939	0.061	0.002	0.871	0.998	Ramon Delgado	0.759	0.746	0.012	0.000	0.737	0.757

Table 3: Original and bias corrected efficiency estimates along side with their 95% bootstrap intervals based on the VRS assumption

Players	VRS	BCvrs	Bias	Std	LB	UB	Players	VRS	BCvrs	Bias	Std	LB	UB
Aaron Krickstein	1.000	0.964	0.036	0.000	0.938	0.998	Igor Andreev	0.934	0.925	0.009	0.000	0.918	0.932
Alberto Berasategui	1.000	0.964	0.036	0.000	0.941	0.999	Olivier Rochus	0.934	0.920	0.014	0.000	0.908	0.932
Alex Corretja	1.000	0.979	0.021	0.000	0.959	0.999	Philipp Kohlschreiber	0.933	0.924	0.008	0.000	0.920	0.932
Alex O'Brien	1.000	0.975	0.025	0.000	0.958	0.998	Sergi Bruguera	0.933	0.918	0.015	0.000	0.904	0.931
Andre Agassi	1.000	0.954	0.046	0.001	0.897	0.999	Juan Martin Del Potro	0.931	0.912	0.019	0.000	0.892	0.930
Andrei Medvedev	1.000	0.983	0.017	0.000	0.972	0.999	Bohdan Ulihrach	0.929	0.912	0.016	0.000	0.899	0.928
Andrew Ilie	1.000	0.953	0.047	0.002	0.899	0.999	Amos Mansdorf	0.928	0.909	0.019	0.000	0.889	0.927
Andy Murray	1.000	0.956	0.044	0.001	0.915	0.999	Karel Novacek	0.928	0.917	0.011	0.000	0.910	0.926
Bernd Karbacher	1.000	0.973	0.027	0.000	0.956	0.999	Jurgen Melzer	0.926	0.915	0.011	0.000	0.907	0.925
Bjorn Phau	1.000	0.966	0.034	0.001	0.926	0.999	Richard Gasquet	0.926	0.912	0.014	0.000	0.899	0.925
Boris Becker	1.000	0.976	0.024	0.000	0.953	0.999	Taylor Dent	0.926	0.911	0.015	0.000	0.900	0.924
Brad Gilbert	1.000	0.966	0.034	0.000	0.943	0.999	Benjamin Becker	0.924	0.903	0.020	0.000	0.879	0.922
Bryan Shelton	1.000	0.955	0.045	0.001	0.909	0.999	Nicolas Lapentti	0.923	0.909	0.014	0.000	0.895	0.922
Carlos Moya	1.000	0.964	0.036	0.001	0.919	0.999	Mario Ancic	0.924	0.906	0.018	0.000	0.891	0.922
Chris Woodruff	1.000	0.956	0.044	0.001	0.916	0.999	Jaime Yzaga	0.923	0.909	0.015	0.000	0.894	0.922
David Ferrer	1.000	0.981	0.019	0.000	0.965	0.999	Emilio Sanchez	0.921	0.909	0.012	0.000	0.900	0.919
David Sanchez	1.000	0.965	0.035	0.001	0.924	0.999	Andrei Cherkasov	0.919	0.897	0.022	0.000	0.872	0.918
David Wheaton	1.000	0.967	0.033	0.000	0.946	0.998	Hicham Arazi	0.919	0.904	0.014	0.000	0.893	0.917
Dmitry Tursunov	1.000	0.978	0.022	0.000	0.966	0.999	Paul Haarhuis	0.916	0.904	0.012	0.000	0.896	0.915
Fabrice Santoro	1.000	0.949	0.051	0.002	0.884	0.999	Albert Montanes	0.916	0.907	0.009	0.000	0.902	0.915
Feliciano Lopez	1.000	0.975	0.025	0.000	0.959	0.999	Kenneth Carlsen	0.916	0.899	0.017	0.000	0.887	0.914
Fernando Verdasco	1.000	0.972	0.028	0.000	0.954	0.999	Olivier Delaitre	0.916	0.899	0.017	0.000	0.882	0.914
Fernando Vicente	1.000	0.973	0.027	0.000	0.955	0.999	Karol Kucera	0.913	0.898	0.014	0.000	0.884	0.912
Francisco Clavet	1.000	0.976	0.024	0.000	0.957	0.999	Daniel Vacek	0.913	0.901	0.012	0.000	0.892	0.911
Gael Monfils	1.000	0.969	0.031	0.000	0.939	0.998	Juan Antonio Marin	0.912	0.889	0.023	0.000	0.850	0.910
Gaston Gaudio	1.000	0.986	0.014	0.000	0.974	0.999	Byron Black	0.911	0.897	0.015	0.000	0.883	0.910
Gianluca Pozzi	1.000	0.964	0.036	0.001	0.939	0.999	Andrea Gaudenzi	0.911	0.899	0.012	0.000	0.891	0.910
Gilles Simon	1.000	0.986	0.014	0.000	0.976	0.999	Filippo Volandri	0.910	0.893	0.017	0.000	0.875	0.908
Goran Ivanisevic	1.000	0.948	0.052	0.002	0.866	0.999	Tomas Berdych	0.909	0.896	0.012	0.000	0.884	0.907
Guillermo Coria	1.000	0.950	0.050	0.002	0.887	0.999	Jose Acasuso	0.908	0.897	0.011	0.000	0.889	0.907
Hendrik Dreekmann	1.000	0.959	0.041	0.001	0.910	0.999	Arnaud Boetsch	0.908	0.896	0.011	0.000	0.887	0.906
Ivan Lendl	1.000	0.951	0.049	0.002	0.899	0.999	Max Mirnyi	0.907	0.892	0.014	0.000	0.880	0.905
Ivo Karlovic	1.000	0.947	0.053	0.003	0.865	0.999	Hernan Gumy	0.906	0.891	0.015	0.000	0.876	0.904
Jan-Michael Gambill	1.000	0.969	0.031	0.000	0.947	0.999	Renzo Furlan	0.905	0.895	0.011	0.000	0.887	0.904
Jeff Tarango	1.000	0.982	0.018	0.000	0.970	0.999	Marc Rosset	0.904	0.885	0.018	0.000	0.868	0.902
Jeremy Chardy	1.000	0.947	0.053	0.002	0.871	0.999	Fernando Meligeni	0.900	0.888	0.012	0.000	0.880	0.899
John Isner	1.000	0.948	0.052	0.002	0.867	0.998	Igor Kunitsyn	0.900	0.885	0.015	0.000	0.871	0.899
Jordi Arrese	1.000	0.965	0.035	0.001	0.934	0.999	Stefan Koubek	0.900	0.885	0.014	0.000	0.871	0.898
Juan Carlos Ferrero	1.000	0.981	0.019	0.000	0.967	0.999	Adrian Voinea	0.899	0.886	0.013	0.000	0.877	0.898
Juan Ignacio Chela	1.000	0.978	0.022	0.000	0.963	0.999	Marcelo Rios	0.898	0.883	0.015	0.000	0.864	0.897
Lleyton Hewitt	1.000	0.957	0.043	0.001	0.911	0.999	Nicolas Massu	0.898	0.882	0.016	0.000	0.867	0.897
Marat Safin	1.000	0.968	0.032	0.000	0.942	0.999	Ernests Gulbis	0.897	0.880	0.017	0.000	0.864	0.896
Marcel Granollers	1.000	0.968	0.032	0.000	0.939	0.999	Karim Alami	0.891	0.877	0.014	0.000	0.866	0.890
Marcos Baghdatis	1.000	0.985	0.015	0.000	0.973	0.999	Jerome Golmard	0.890	0.879	0.011	0.000	0.871	0.889
Mark Philippoussis	1.000	0.969	0.031	0.000	0.942	0.998	Brett Steven	0.889	0.876	0.013	0.000	0.867	0.888
Michael Chang	1.000	0.961	0.039	0.001	0.919	0.999	Raemon Sluiter	0.886	0.871	0.015	0.000	0.854	0.884
Mikael Tillstrom	1.000	0.967	0.033	0.000	0.940	0.999	Andrei Pavel	0.882	0.868	0.013	0.000	0.857	0.881
Novak Djokovic	1.000	0.978	0.022	0.000	0.957	0.999	Mariano Puerta	0.878	0.865	0.013	0.000	0.854	0.877
Patrick McEnroe	1.000	0.962	0.038	0.001	0.930	0.999	Kristof Vliegen	0.877	0.859	0.018	0.000	0.840	0.876
Paul Goldstein	1.000	0.948	0.052	0.002	0.869	0.999	Christophe Rochus	0.876	0.861	0.015	0.000	0.847	0.875

Pete Sampras	1.000	0.952	0.048	0.001	0.896	0.999	Janko Tipsarevic	0.869	0.855	0.014	0.000	0.844	0.868
Rafael Nadal	1.000	0.963	0.037	0.001	0.928	0.999	Paradorn Srichaphan	0.866	0.854	0.012	0.000	0.846	0.865
Richard Fromberg	1.000	0.968	0.032	0.000	0.950	0.999	Sebastien Lareau	0.864	0.849	0.015	0.000	0.836	0.863
Richard Krajicek	1.000	0.960	0.040	0.001	0.921	0.998	Florian Mayer	0.863	0.853	0.010	0.000	0.847	0.862
Roger Federer	1.000	0.948	0.052	0.002	0.870	0.998	Marcelo Filippini	0.855	0.841	0.014	0.000	0.826	0.854
Sandon Stolle	1.000	0.977	0.023	0.000	0.954	0.999	Filip Dewulf	0.853	0.837	0.015	0.000	0.821	0.851
Scott Draper	1.000	0.973	0.027	0.000	0.951	0.999	Shuzo Matsuoka	0.849	0.832	0.017	0.000	0.813	0.848
Thomaz Bellucci	1.000	0.960	0.040	0.001	0.916	0.999	Carl-Uwe Steeb	0.845	0.831	0.014	0.000	0.820	0.844
Todd Martin	1.000	0.959	0.041	0.001	0.926	0.998	Christian Ruud	0.844	0.831	0.013	0.000	0.817	0.843
Tomas Carbonell	1.000	0.970	0.030	0.000	0.950	0.999	Yen-Hsun Lu	0.840	0.823	0.016	0.000	0.807	0.839
Vincent Spadea	1.000	0.967	0.033	0.001	0.926	0.999	Jakob Hlasek	0.838	0.826	0.012	0.000	0.816	0.837
Wayne Arthurs	1.000	0.949	0.051	0.002	0.868	0.999	Hyung-Taik Lee	0.808	0.799	0.008	0.000	0.795	0.807
Wayne Ferreira	1.000	0.975	0.025	0.000	0.953	0.999	Luis Horna	0.798	0.785	0.013	0.000	0.772	0.797
Yevgeny Kafelnikov	1.000	0.971	0.029	0.000	0.937	0.999	Ramon Delgado	0.769	0.750	0.019	0.000	0.727	0.768

Additionally to table 3, table 4 provides us with the descriptive statistics of the obtained efficiency estimates under the VRS and CRS case both for the original and biased corrected estimates along side with their 95% bootstrapped confidence intervals. The descriptive statistics reveal that biased corrected efficiency estimates are lower compared to the original efficiency scores both for the CRS and the VRS case.

The VRS scores are also higher compared to the CRS scores since the number of tennis players reported to be efficient are greater for the VRS case. The standard deviations values presents that the efficiency variations among the 229 tennis players are not so large indicating a high competitive environment among the professional tennis players.

Figure 2 presents the density plots of the estimated efficiency scores both for the CRS and VRS scales. The black lines indicate the CRS efficiency estimates (original-solid line and biased corrected-dashed line) whereas the blue lines indicate the VRS efficiency estimates (original-dotted line and biased corrected-dash dotted line). As can be observed for the distribution of the original CRS efficiency estimates appears to be platykurtic (i.e. having a negative excess kurtosis) indicating a lower, wider peak around tennis players' mean efficiency estimates.

Table 4: Descriptive statistics of professional tennis players efficiency scores

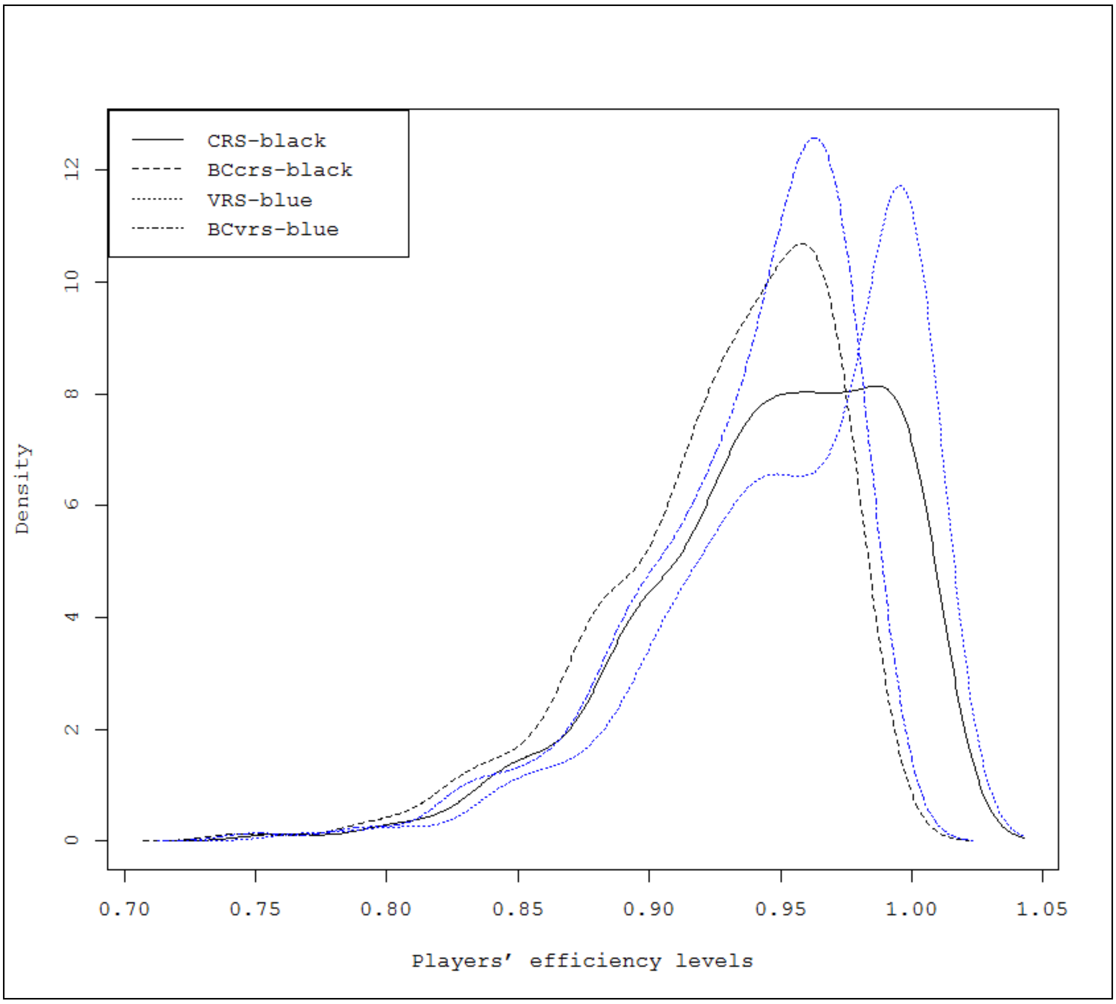
Constant returns to scale model-CRS						
	CRS	*BCcrs*	*Bias*	*Std*	*LB*	*UB*
max	1.000	0.983	0.061	0.003	0.974	0.999
min	0.759	0.746	0.007	0.000	0.737	0.757
mean	0.944	0.927	0.018	0.000	0.913	0.943
std	0.047	0.043	0.009	0.000	0.042	0.047
Variable returns to scale model-VRS						
	VRS	*BCvrs*	*Bias*	*Std*	*LB*	*UB*
max	1.000	0.986	0.053	0.003	0.976	0.999
min	0.769	0.750	0.008	0.000	0.727	0.768
mean	0.956	0.935	0.021	0.000	0.915	0.954
std	0.046	0.042	0.011	0.000	0.041	0.046

But the distribution of the biased corrected efficiency estimates for the CRS case appeared to be leptokurtic having a sharp peak around tennis players' mean efficiency estimates. Moreover, under the assumption of VRS it appears that the distribution of tennis players' efficiency has a tendency towards a polarization of twin peaks (i.e. of tennis players having an efficiency score of 0.95 and of those having an efficiency score equal to 1). However the tendency of the twin-peakedness effect disappears when looking at the distribution for the biased corrected estimates under the VRS case which appear to be leptokurtic.

Finally, following the methodology proposed by Simar and Wilson (2002) we test whether the CRS or VRS assumption is better suited for our efficiency analysis. In our application we have one input and nine outputs and we obtained for this test a p-value of 0.3187> 0.05 (with B=2000) hence, we can not reject the null hypothesis of

CRS. Therefore, the results which are more suitable in our analysis are the ones based on the CCR model assuming constant returns to scale[12].

Figure 2: Density plots of the original and biased corrected efficiency scores for the CRS and the VRS case.



Therefore the thirty nine professional players who are efficient under the CRS assumption (efficiency $=1$) in alphabetical order are: Aaron Krickstein, Alberto Berasategui, Alex O'Brien, Andre Agassi, Andrew Ilie, Andy Murray, Bernd Karbacher, Boris Becker, Brad Gilbert, Bryan Shelton, David Wheaton, Dmitry Tursunov, Fernando Verdasco, Fernando Vicente, Gael Monfils, Gianluca Pozzi,

[12] The analytical results for the CRS and VRS case are available upon request.

Guillermo Coria, Hendrik Dreekmann, Ivan Lendl, Ivo Karlovic, Jan-Michael Gambill, John Isner, Lleyton Hewitt, Marat Safin, Mark Philippoussis, Michael Chang, Mikael Tillstrom, Novak Djokovic, Patrick McEnroe, Paul Goldstein, Pete Sampras, Rafael Nadal, Richard Fromberg, Richard Krajicek, Roger Federer, Scott Draper, Todd Martin, Tomas Carbonell and Wayne Arthurs.

5. Conclusions

This paper evaluates the performance of 229 professional tennis players using career data from the ATP database. By using a sports production framework it provides unified performance indicators incorporating nine performance measures into a single efficiency indicator. Moreover, by using the DEA methodology the proposed efficiency indicator measures tennis players' career performance by applying an input –output setting.

Therefore the career matches along side with nine different output measures are modelled accordingly under the two basic economic assumptions (i.e. constant and variable returns to scale)[13]. Moreover, bootstrap techniques have been applied for bias correction of the calculated efficiency estimates and for the construction of 95% confidence intervals of the efficiency estimates.

In addition a bootstrap test has been used indicating the appropriateness of the CRS model when analysing the obtained efficiency scores. The results reveal that 39 tennis players appear to be efficient. Finally, the results reveal a high competitive environment between the 229 professional tennis players reflected on their relative small differences among the estimated efficiencies.

[13] However it must be highlighted that other essential factors (external to the proposed sports production function) like coaching provision and coaching development (Kellett, 1999), countries' elite sport policies and systems (Sotiriadou and Shilbury 2009, De Bosscher et al. 2012) and countries mechanisms of talent identification at a young age (Brouwers, De Bosscher & Sotiriadou, 2012) can influence indirectly tennis players' career efficiency levels.

APPENDIX A

Appendix A.1

The 229 professional tennis players in alphabetical order as extracted from the Association of Tennis Professional (ATP)

Aaron Krickstein, Adrian Voinea, Agustin Calleri, Albert Costa, Albert Montanes, Albert Portas, Alberto Berasategui, Alberto Martin, Alex Corretja, Alex O'Brien, Alexander Volkov, Amos Mansdorf, Andre Agassi, Andrea Gaudenzi, Andreas Seppi, Andrei Cherkasov, Andrei Chesnokov, Andrei Medvedev, Andrei Olhovskiy, Andrei Pavel, Andrew Ilie, Andy Murray, Andy Roddick, Antony Dupuis, Arnaud Boetsch, Arnaud Clement, Benjamin Becker, Bernd Karbacher, Bjorn Phau, Bohdan Ulihrach, Boris Becker, Brad Gilbert, Brett Steven, Bryan Shelton, Byron Black, Carlos Costa, Carlos Moya, Carl-Uwe Steeb, Cedric Pioline, Chris Woodruff, Christian Ruud, Christophe Rochus, Daniel Nestor, Daniel Vacek, David Ferrer, David Nalbandian, David Prinosil, David Sanchez, David Wheaton, Davide Sanguinetti, Dmitry Tursunov, Dominik Hrbaty, Emilio Sanchez, Ernests Gulbis, Fabio Fognini, Fabrice Santoro, Feliciano Lopez, Felix Mantilla, Fernando Gonzalez, Fernando Meligeni, Fernando Verdasco, Fernando Vicente, Filip Dewulf, Filippo Volandri, Florent Serra, Florian Mayer, Francisco Clavet, Franco Squillari, Gael Monfils, Galo Blanco, Gaston Gaudio, Gianluca Pozzi, Gilbert Schaller, Gilles Muller, Gilles Simon, Goran Ivanisevic, Greg Rusedski, Guillaume Raoux, Guillermo Canas, Guillermo Coria, Guillermo Garcia-Lopez, Gustavo Kuerten, Guy Forget, Hendrik Dreekmann, Hernan Gumy, Hicham Arazi, Hyung-Taik Lee, Igor Andreev, Igor Kunitsyn, Ivan Lendl, Ivan Ljubicic, Ivo Karlovic, Jacco Eltingh, Jaime Yzaga, Jakob Hlasek, James Blake, Jan Siemerink Janko Tipsarevic, Jan-Michael Gambill, Jarkko Nieminen, Jason Stoltenberg, Javier Frana, Javier Sanchez, Jeff Tarango, Jeremy Chardy, Jerome Golmard, Jim Courier, Jiri Novak, John Isner, Jonas Bjorkman, Jonathan Stark, Jordi Arrese, Jordi Burillo, Jose Acasuso, Jo-Wilfried Tsonga, Juan Antonio Marin, Juan Carlos Ferrero, Juan Ignacio Chela, Juan Martin Del Potro, Juan Monaco, Julien Benneteau, Jurgen Melzer, Justin Gimelstob, Karel Novacek, Karim Alami, Karol Kucera, Kenneth Carlsen, Kristof Vliegen, Lars Burgsmuller, Lleyton Hewitt, Luis Horna, Magnus Gustafsson, Magnus Larsson, Magnus Norman, MaliVai Washington, Marat Safin, Marc Rosset, Marcel Granollers, Marcelo Filippini, Marcelo Rios, Marc-Kevin Goellner, Marcos Baghdatis, Marcos Ondruska, Mardy Fish, Mariano Puerta, Mariano Zabaleta, Marin Cilic, Mario Ancic, Mark Philippoussis, Mark Woodforde, Martin Damm, Max Mirnyi, Michael Chang, Michael Llodra, Michael Stich, Mikael Tillstrom, Mikhail Youzhny, Nicklas Kulti, Nicolas Almagro, Nicolas Escude, Nicolas Kiefer, Nicolas Lapentti, Nicolas Mahut, Nicolas Massu, Nikolay Davydenko, Novak Djokovic, Olivier Delaitre, Olivier Rochus, Omar Camporese, Paradorn Srichaphan, Patrick McEnroe, Patrick Rafter, Paul Goldstein, Paul Haarhuis, Paul-Henri Mathieu, Pete Sampras, Petr Korda, Philipp Kohlschreiber, Potito Starace, Radek Stepanek, Raemon Sluiter, Rafael Nadal, Rainer Schuettler, Ramon Delgado, Renzo Furlan, Richard Fromberg, Richard Gasquet, Richard Krajicek, Richey Reneberg, Robby Ginepri, Robin Soderling, Roger Federer, Ronald Agenor, Sam Querrey, Sandon Stolle, Sargis Sargsian, Scott Draper, Sebastien Grosjean, Sebastien Lareau, Sergi Bruguera, Shuzo Matsuoka, Sjeng Schalken, Slava Dosedel, Stanislas Wawrinka, Stefan Edberg, Stefan Koubek, Stefano Pescosolido, Taylor Dent, Thomas Enqvist, Thomas Johansson, Thomas Muster, Thomaz Bellucci, Tim Henman, Todd Martin, Todd Woodbridge, Tomas Berdych, Tomas Carbonell, Tommy Haas, Tommy Robredo, Victor Hanescu, Viktor Troicki, Vincent Spadea, Wally Masur, Wayne Arthurs, Wayne Ferreira, Xavier Malisse, Yen-Hsun Lu, Yevgeny Kafelnikov and Younes El Aynaoui.

Appendix A.2

A synoptic representation of Simar and Wilson's (1998, 2000a, 2000b) bootstrap algorithm

In order to implement the homogenous bootstrap algorithm for a set of bootstrap estimates $\left\{ \hat{\lambda}_b^*(x,y) \middle| b=1,...,B \right\}$ for a given fixed point (x,y) the following eight steps must be carried out:

1. From the original data set we compute $\hat{\lambda}_{DEA}$.
2. Then we apply the "rule of thump" (Silverman 1986, p.45-48) to obtain the bandwidth parameter h.
3. We generate $\beta_1^*,...,\beta_n^*$ by drawing with replacement from the set $\left\{ \hat{\lambda}_1,...,\hat{\lambda}_n, \left(2-\hat{\lambda}_1\right),...,\left(2-\hat{\lambda}_n\right) \right\}$.
4. Then we draw $\varepsilon_i^*, i=1,...,n$ independently from the kernel function $K(.)$ and compute $\beta_i^{**} = \beta_i^* + h\varepsilon_i^*$ for each $i=1,...,n$.
5. For each $i=1,...,n$ we compute β_i^{***} as: $\beta_i^{***} = \bar{\beta}^* + \frac{\beta_i^{**} - \bar{\beta}^*}{\left(1+h^2\sigma_k^2\sigma_\beta^2\right)^{1/2}}$,

   where $\bar{\beta}^* = \sum_{i=1}^{n} \beta_i^* / n$, $\sigma_\beta^2 = \sum_{i=1}^{n} \left(\beta_i^* - \bar{\beta}^* \right) / n$ and σ_k^2 is the variance of the probability density function used for the kernel function. In addition λ_i^* can then be computed as: $\lambda_i^* = \begin{cases} 2-\beta_i^{***} \forall \beta_i^{***} < 1 \\ \beta_i^{***} \text{ otherwise} \end{cases}$.

6. The bootstrap sample is created as: $X_n^* = \left\{ \left(x_i^*, y_i \right) \middle| i = 1,...,n \right\}$ where $x_i^* = \lambda_i^* \hat{x}^{\partial}(y_i) = \lambda_i^* \hat{\lambda}_i^{-1} x_i$.

7. We compute the DEA efficiency estimates $\hat{\lambda}_i^*(x_i, y_i)$ for each of the original sample observations using the reference set X_n^* in order to obtain a set of bootstrap estimates.

8. Finally, we repeat steps 3 to 7 B times (at least 2000 times) to obtain a set of bootstrap estimates $\left\{ \hat{\lambda}_b^*(x, y) \middle| b = 1,...,B \right\}$.

References

Atkinson, S.E., Stanley, L. R., & Tschirart, J. (1988). Revenue sharing as an incentive in an agency problem: An example from the National Football League. *Rand Journal of Economics*, 19(1), 27-43.

Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* 30(9), 1078– 1092.

Barros, C.P. (2003). Incentive regulation and efficiency in sport organisational training activities. *Sports Management Review* 6(1), 33-52.

Barros, C. P., & Leach, S. (2006a). Analyzing the performance of the English F.A. Premier league with an econometric frontier model. *Journal of Sports Economics*, 7(4), 391-407.

Barros, C. P., & Leach, S. (2006b). Performance evaluation of the English premier league with data envelopment analysis. *Applied Economics*, 38(12), 1449-1458.

Barros, C. P., & Leach, S. (2007). Technical efficiency in the English football association premier league with a stochastic cost frontier. *Applied Economics Letters*, 14 (10), 731-741.

Barros, C.P., & Garcia-del-Barrio, P. (2008). Efficiency measurement of the English Football Premier League with a Random Frontier Model. *Economic Modelling*, 25(5), 994–1002.

Barros, C.P., Assaf, A., & Sá-Earp, F. (2010). Brazilian football league technical efficiency: A Simar and Wilson Approach. *Journal of Sports Economics*, 11(6), 641-651.

Bashtannyk, D.M., & Hyndman, R.J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational statistics and data analysis*, 36(3), 279-298.

Boulier, B. L., & Stekler, H. O. (1999). Are sports seedings good predictors? An evaluation. *International Journal of Forecasting*, 15(1), 83–91.

Brouwers, J., De Bosscher, V., & Sotiriadou, P. (2012) An examination of the importance of performances in youth and junior competition as an indicator of later success in tennis. *Sports Management Review*, 15(4), 461-475.

Carmichael, F., & Thomas, D. (1995). Production and efficiency in team sports: An investigation of rugby league football. *Applied Economics*, 27(9), 859-869.

Charnes, A., Cooper, W.W., & Rhodes, E.L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444.

Clarke, S. R., & Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research*, 7(6), 585–594.

Coate, D., & Robbins, D. (2001). The tournament careers of top-ranked men and women tennis professionals: Are the gentlemen more committed than the ladies? *Journal of Labor Research*, 22(1), 185-193.

Coelli, T.J. , Rao, D.S.P., O'Donnell C.J., & Battese, G.E. (2005). *An introduction to efficiency and productivity analysis.* NY, USA: Springer.

Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis*. New York: Springer Science.

Dawson, P., Dobson, S., & Gerrard, B. (2000a). Stochastic frontiers and the temporal structure of managerial efficiency in English soccer. *Journal of Sport Economics*, 1(4), 341-362.

Dawson, P., Dobson, S., & Gerrard, B. (2000b). Estimating coaching efficiency in the professional team sports: Evidence from English association football. *Scottish Journal of Political Economy*, 47(4), 399- 421.

De Bosscher V., De Knopa P., van Bottenburg M., Shibli S., & Binghamd, J. (2012). Explaining international sporting success: An international comparison of elite sport systems and policies in six countries. *Sports Management Review* 12(3), 113-136.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrics*, 19(3), 273–292.

del Corral, J., & Prieto-Rodríguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*, 26(3), 551–563.

Färe R., & Lovell, C.A.K. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1), 150-162.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A*, 120(3), 253–281.

Fizel, J.L., & D'Itri, M.P. (1996). Estimating managerial efficiency: the case of college basketball coaches. *Journal of Sport Management*, 10(4), 435–445.

Fizel, J.L. and D'Itri, M.P. (1997). Managerial efficiency, managerial succession and organizational performance. *Managerial and Decision Economic*, 18(1), 295–308.

Førsund, F.R., & Sarafoglou, N. (2002). On the origins of Data Envelopment Analysis. *Journal of the Productivity Analysis*, 17(1/2), 23-40.

Førsund, F.R., Kittelsen, S.A.C., & Krivonozhko, V.E. (2009). Farrell revisited–Visualizing properties of DEA production frontiers. *Journal of the Operational Research Society*, 60(11), 1535-1545.

Frick, B., & Simmons, R. (2008). The impact of managerial quality on organizational performance: Evidence from German soccer. *Managerial and Decision Economics*, 29(7), 593-600.

Haas, D.J. (2003a). Productive efficiency of English football teams-A data envelopment analysis approach. *Managerial and Decision Economics*, 24(5), 403-410.

Haas, D.J. (2003b). Technical efficiency in the major league soccer. *Journal of Sport Economics*, 4(3), 203-215.

Hyndman, R.J., Bashtannyk, D.M., & Grunwald, G.K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5(4), 315-336.

Hyndman, R.J., & Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *Journal of Nonparametric Statistics*, 14(3), 259-278.

Jones, J.C.H. (1969). The Economics of the National Hockey League. *Canadian Journal of Economics*, 2(1), 1-20.

Kellett, P. (1999). Organisational leadership: Lessons from professional coaches. *Sports Management Review*, 2(2), 150–171.

Klaassen, F., & Magnus, J. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148(2), 257–267.

Lee, Y.H., & Berri, D. (2008). A re-examination of production functions and efficiency estimates for the national basketball association. *Scottish Journal of Political Economy*, 55(1), 51-66.

McHale, I., & Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2), 619–630.

Neale, W. (1964). The peculiar economics of professional sports. *Quarterly Journal of Economics*, 78(1), 1-14.

O' Donnell, C.J. (2012). Econometric estimation of distance functions and associated measures of productivity and efficiency change. *Journal of Productivity Analysis*, DOI 10.1007/s11123-012-0311-1.

Porter, P. & Scully, G.W. (1982). Measuring managerial efficiency: The case of baseball. *Southern Economic Journal*, 48(3), 642-650.

Rámon, N., Ruiz, J.L., & Sirvent, I. (2012). Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players. *Expert Systems with Applications*, 39(5), 4882–4889.

Rohm, A.J., Chatterjee, S., & Habibullah, M. (2004) Strategic measure of competitiveness for ranked data. *Managerial and Decision Economics*, 25(2), 103–108.

Rottenberg, S. (1956). The baseball player's labor-market. *Journal of Political Economy*, 64(3), 242-258.

Scheibehenne, B., & Broder, A. (2007). PredictingWimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, 23(3), 415–426.

Schofield, J. A. (1988). Production functions in the sports industry: An empirical analysis of professional cricket. *Applied Economics*, 20(2), 177-193.

Scully, G.W. (1974). Pay and performance in major league baseball. *American Economic Review*, 64(6), 915-930.

Shephard, R.W. (1970). *Theory of cost and production functions*. Princeton: Princeton University Press.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simar, L. & Wilson, P.W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science*. 44(1), 49–61.

Simar, L. & Wilson, P.W. (2000a). A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics*, 27(6), 779–802.

Simar, L. & Wilson, P.W. (2000b). Statistical inference in nonparametric frontier models: the state of the art. *Journal of Productivity Analysis*, 13(1), 49–78.

Simar, L. & Wilson, P.W. (2002). Nonparametric tests of returns to scale. *European Journal of Operational Research*, 139(1), 115– 132.

Simar, L. & Wilson, P.W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64.

Sloane, P.J. (1969). The labour market in professional football. *British Journal of Industrial Relations*, 7(2), 181-199.

Sloane, P.J. (1971). The economics of professional football: The football club as a utility maximiser. *Scottish Journal of Political Economy*, 17(2), 121-146.

Sloane, P.J. (1976). Restriction of competition in professional team sports. *Bulleting of Economic Research*, 28(1), 3-22.

Sotiriadou, K., & Shilbury, D. (2009). Australian Elite Athlete Development: An Organisational Perspective. *Sports Management Review*, 12(3), 137-148.

Wozniak, D. (2012). Gender differences in a market with relative performance feedback: Professional tennis players. *Journal of Economic Behavior and Organization*, 83(1), 158– 171.

Zak, T. A., Huang, C. J., & Siegfried, J.J. (1979). Production efficiency: The case of professional basketball. *Journal of Business*, 52(3), 379-392.

Zech, C. E. (1981). An empirical estimation of a production function: The case of Major League Baseball. *American Economist*, 25(2) ,19-23.