International Journal of Sport, Exercise & Training Sciences
e-ISSN 2149-8229 Vol 5, Issue 4, 201–213, (2019) *Original Article*
http://dergipark.gov.tr/useeabd

Partial Least Squares-Structural Equation Modeling (PLS-SEM) Analysis of Team Success Using R

Mehmet TÜREGÜN[1]

Abstract

Aim: A combination of football becoming highly commercialized, technological advances made, and increasing amounts of data becoming available has made it possible for researchers to conduct statistical analyses of the various aspects of the game with an ultimate focus on determining the key factors for team success.

Methods: This quasi-experimental study used an ex-post facto design to develop a model for team success. The sample consisted of 18 teams which played 306 matches in a 9-month long association football league format. A PLS-SEM path analysis was conducted using 11 latent variables.

Results: Findings yielded a substantial overall model fit (GoF R^2=0.811) for the measurement and structural models. The latent variables (LVs) of *offense* (β= 0.630, p< .001) and *defense* (β= 0.489, p< .001) had statistically significant effects on the LV of *success*. The exogenous LVs *offense* and *defense* predicted 79.9% of the variability of the LV *success* and its manifest variables.

Conclusion: The defensive ability of a team seemed just as important as the offensive ability for team success in football. This particular conclusion is well aligned with the outcome of various studies conducted by other researchers. For instance, Hughes & Churchill (2004) stated that in their study it appeared that defensive ability of teams to control the opposing team's movements had a significant effect on team success.

Keywords
Team success,
modeling,
path analysis,
football,

Article Info
Received: 03.10.2019
Accepted: 12.12.2019
Online Published: 30.12.2019

DOI:10.18826/useeabd.628653

INTRODUCTION

Football, or known as soccer in the United States, is a free-flowing game, with the play naturally going back and forth from defense to offense (Wade, 1996). The game of football is a team sport played by two teams consisting of eleven players on each team. As football has become a highly commercialized industry, the necessity of success has utmost importance in football, and this is evidenced by the presence in the literature of several research studies that focused on identifying indicators of success (Dufour, Phillips, and Ernwein, 2017; James, Jones, and Mellalieu, 2004). Over the years, there have been many rating systems designed to assess current levels of skill and success for club and national teams. A Soccer Power Index (SPI) established by Entertainment and Sports Programming Network (ESPN), the Federation of the International Football Association (FIFA) ranking system and the Elo rating system, which was devised by Arpad Elo who was a Hungarian-American physicist, are just a few examples of index of success. As opposed to other rating systems, SPI, designed as a predictive system, aims to project the best possible representation of team success looking forward. A detailed explanation of ESPN's SPI, which is presented by the ESPN staff, can be obtained from the web-site located at https://www.espn.com/soccer/news/story/_/id/1873765.

The primary objective in football is to win the match, or at least not to lose. In order to reach that objective, a team must outscore the opposition. Hence, several researchers reported that scoring goals was an important determinant of success (Carling, Le Gall, McCall, Nédélec, and Dupont, 2015; Dufour et al., 2017; Hughes & Franks, 2005; Jones, James, and Mellalieu, 2004). Consequently, some of the more important parameters of index of success defined and used by SPI are based on goals scored and conceded at home and away, not necessarily wins, losses and draws. It is reported by the ESPN staff that since the 1998-1999 season, the English Premier League (EPL) teams having a better goal differential in league play, but fewer league points, have a record of 179 wins, 138 losses and 130 draws. In other top leagues, such as Spanish La Liga and Italian Serie A, similar trends can be observed. SPI is, in part, based on the correlations between the future success and scoring margins, whereas other systems, such as the FIFA rankings and the Elo ratings, are based on wins and losses. According to the ESPN staff, the outcome of SPI consists of offensive and defensive ratings, which

[1]Barry Üniversity, United States, mturegun@barry.edu **ORCID ID:** 0000-0001-8264-6996

are defined as the number of goals a team would be expected to score or concede against a league-average team at a neutral site. Hence, the ESPN staff claims that SPI is better at projecting future success than the other similar systems and methods.

In this paper, I offer a simple index of success model using Partial Least Square-Structural Equation Modeling (PLS-SEM) implemented in statistical computing platform R version 3.4.4 (R Core Team, 2018) package *plspm*, which was developed by Sanchez, Trinchera and Russolillo (2017). I used Wade's (1996) concept of team success in football and employed a slightly modified version of the simple index of success model devised by Sanchez (2013), who illustrated the basic concepts of PLS-SEM by using data from the Spanish La Liga for the 2008-2009 season. By using the context of the Turkish Super League in the 2016-2017 season, I aimed to demonstrate the basics of how to implement a PLS-SEM path analysis technique. To my knowledge, these two studies are the only ones implementing PSL-SEM techniques to investigate and identify indicators of team success in the context of professional football.

Review of the Related Literature: The highly competitive and commercialized nature of football at the elite club and national levels have resulted in an increasing need for innovate tactics and strategies to ensure team success. Hence, several previous research studies have focused on identifying various indicators that determine team success in football (Dufour et al., 2017; Hughes & Bartlett, 2002; Hughes & Churchill, 2004; Lago-Peñas & Dellal, 2010). As an overall combination of attacking and defensive styles of play improves the chances of team success, many elite teams have been reported to increasingly use compact defensive mindsets while implementing strong attacking tactics the same time (Fernandez-Navarro, Fraduab, Zubillagac, Forda, and McRoberta, 2016; Tenga, Holme, Ronglan, and Bahr, 2010). Hence, the team success may be hinged on both scoring and not conceding a goal or goals. A link between the number of goals scored at home and away, as well as the number of goals conceded at home and away, and success in football was suggested by several researchers (Carling et al., 2015; Dufour et al., 2017).

Although the link between the goals scored and/or conceded and success is well established, the relationship between team success and other parameters, such as shots on goal, cards received and passes completed does not seem to be as clearly identified. Based on an analysis of three consecutive FIFA World Cup tournaments, Castellano, Casamichana, and Lago (2012) reported that shots on goal, as an attacking play variable, had a high discriminatory power among the winning, drawing and losing teams. However, they reported that this finding was not consistent across all three of the tournaments included in their study. Similarly, in another a study focusing on FIFA 2014 World Cup tournament, Dufour et al. (2017) reported that qualified and non-qualified teams did not differ in the number of shots on goal with a Cohen's d effect size of .15. On the other hand, they defined and used shooting efficiency as an indicator and reported that it had an impact on winning or qualifying during the FIFA 2014 World Cup. Moreover, according to Hughes & Bartlett (2002), success in football can be divided into passes, tackles and shots. For instance, overall number of passes, pass attempts and overall number of accurate passes were reported as important factors in achieving better results (Lago-Peñas & Dellal, 2010; Saito, Yoshimura, and Ogiwara, 2013). On the other hand, types of passes were not seen as a performance indicator (Scoulding, James, and Taylor, 2004).

Additionally, both tactical and technical factors can determine team success. Ball possession has been widely reported as one of the most important tactical determinants of team success in football (Hughes & Churchill, 2004; Tenga & Sigmundstad, 2011). In general, ball possession seemed to be influenced by situational variables, such as game outcome, game location, the type of competition, and the quality of opponent (Lago-Peñas & Dellal, 2010). In general, successful teams had more ball possession compared to unsuccessful teams. However, there was no difference reported between successful and non-successful teams in terms of ball possession when winning (James et al., 2004). This is evidenced by the fact that having ball possession for a long time does not necessarily guarantee or lead to goal scoring. On the contrary, it is rather possible that ball possession might provide an opponent with time necessary to organize in a better defensive formation and leave less space into which for the attacking team to play. An analysis of goal scoring organizations has been used to detect game patterns in order to differentiate successful teams from non-successful teams. These patterns can be grouped into offensive and defensive patterns. According to Hughes & Churchill (2004), it appeared that defensive ability of teams to control the opposing team's movements had a great effect on team success.

As another indicator of team success, the final league standings have been used by several research studies to distinguish between successful and non-successful teams. Among other indicators, such as ball possession, tackle outcomes, nature of shots, and nature of passing, successful teams were significantly different than non-successful teams in goals scored and conceded at home and away (Lago-Peñas & Dellal, 2010). Additionally, findings of the analysis reported by Szwarc (2007) suggested that for the highest level of competitions the most crucial issue was to have a strong offense and attack to score a goal. After that is achieved, the teams seemed to try to keep the score by using simple technical and tactical actions in defense. Furthermore, it was reported by several researchers that teams aimed to achieve a two-fold outcome. This two-fold outcome, which is considered to be one of the most valid indicators to determine team success, is described as scoring goals and preventing the opposing team from scoring goals (Armatas, Yiannakos, and Sileloglou, 2007; Hughes & Franks, 2005; Kempe, Vogelbein, Memmert, and Nopp, 2014; Tenga et al., 2010; Tenga & Sigmundstad, 2011). In addition to these researchers, Lanham (2005) argued that scoring and conceding goals were important determinants of team success and performance in football both at national and club level competitions.

As goals scored and conceded at home and away were reported to be strong indicators of offensive and defensive characteristics in the literature, scoring and conceding goals were treated as main indicators for team success in this study. Likewise, according to Wade (1996), team success in football is acquired through effectively performing the three fundamental phases of play. These three phases are attack (or offense), defense and preparation. The third phase of play, the preparation phase is based on passing sequences and ball possession, which are, in turn, can be molded by a specific coaching and/or tactical philosophy and style of play. Having shifted the attention on the first two phases of play, one might devise a model based on Wade's theory. Hence, a basic theoretical model for team success can summarized as follows: The better the quality of the attack or offensive schemes, as well as the quality of the defensive strategies and tactics are, the more success the teams have.

Inspired by the legendary Dutch footballer Johan Cruyff's phrase playing football is very simple, but playing simple football is not, I propose a basic model for team success in football. The model I used is a slightly modified version of the simple index of success model devised by Sanchez (2013), who illustrated the basic concepts of PLS-SEM by using data from the Spanish La Liga for the 2008-2009 season. Figure 2 depicts the model, which is partially based on Wade's (1996) theory of team success in football.

Furthermore, this basic theory of team success in football may involve two hypotheses. First, I hypothesize that if a team improves its offense by attacking more, it should be more successful and therefore, win more matches. Secondly, I hypothesize that if a team improves its defense, it should also be more successful, or at least it should avoid losing matches. This can be formulated in the form of the following multiple regression equation

$$\text{success} = b_1\,(\text{offense}) + b_2\,(\text{defense})$$

where b_1 and b_2 are the model coefficients that are greater than zero for the latent variables of offense and defense, respectively.

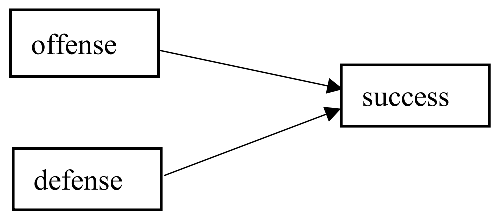

Figure 2. An index of team success model

METHOD

As a methodological framework, statistical modeling is based on developing an understanding and explaining variation with the ultimate goal of estimating parameters in a specific model that leads to the best fit to the data (Crawley, 2007). In achieving the objective of estimating parameters within the statistical modeling framework, the choice of appropriate statistical analysis plays an important role. In turn, the choice of the appropriate statistical analysis is dependent upon the independent variables (IVs), dependent variables (DVs), types of measures, types of variables, factors, and/or levels of data available to the researcher.

As suggested by Hoyle (1995), I used Structural Equation Modeling (SEM) to test hypotheses about the relationships among observed and latent variables. Observed variables are sometimes referred to as indicator variables or manifest variables (MVs), as well. Latent variables (LVs), which cannot be measured directly, are also referred to as unobserved variables, such as satisfaction, self-confidence, motivation, depression, intelligence, and success. In general, SEM consists of two parts known as measurement and structural models. Specifying a model based on theory and testing the initially specified model may not be meaningful unless the measurement model holds. Hence, researchers often test the measurement model before the structural model.

Participants/Sample

The sample consisted of the 306 home and away matches played by 18 teams competed in the Turkish Super League (TSL) during the 2016-2017 season, which spanned a 9-month time frame from August to May. The TSL uses a system of competition that is known as the traditional league system, in which each team plays one home and one away match against the other teams, with three, one and no points awarded for a win, draw and a defeat, respectively. The analysis used here is based on the data collected from these 306 matches played. Figure 1 displays the geographic locations of the 18 teams competed in the TSL during the 2016-2017 season. The two largest cities In Turkey, Istanbul and Ankara were the only two cities represented by multiple teams. Istanbul had 5 and Ankara had 2 teams participating in the competition. The data for our analysis was obtained from the website located at https://footystats.org/turkey/super-lig/2016-2017/home-away-league-table.

Figure 1. The geographic locations of the professional football teams competed in the TSL during the 2016-2017 season.

Research Design and the Variables

This quasi-experimental study used an ex-post facto design in order to investigate specific performance indicators of success for a group of football teams competed in the TSL during the 2016-2017 season. The data consisted of 11 variables measured on 18 teams participated in a 9-month long competition played in the traditional association football league format. The variables can be divided into three categories of Offense Related Variables (ORVs), Defense Related Variables (DRVs) and Success Related Variables (SRVs). The ORVs are total number of goals scored at home (tgsh), total number of goals scored away (tgsa), percentage of matches with goals scored at home (pmsh), percentage of matches with goals scored away (pmsa), and percentage of matches with goals scored away (pmsa). The DRVs are total number of against goals at home (tagh), total number of against goals away (taga), percentage of matches with no against goals at home (pmnagh), and percentage of matches with no against goals away (pmnaga). The SRVs are total number of wins at home (twh), total number of wins away (twa), and total points at the end of season (tpts). Table 1 summarizes the latent variables (LVs), manifest variables (MVs), and their descriptions.

Table 1. Latent variables, manifest variables (i.e., indicators) and their descriptions

Variable Type	Variable Nature	Description
LV	**Offense**	
MV	tgsh	total number of goals scored at home
MV	Tgsa	total number of goals scored away
MV	Pmsh	percentage of matches with goals scored at home
MV	Pmsa	percentage of matches with goals scored away
LV	**Defense**	
MV	Tagh	total number of goals conceded at home
MV	Taga	total number of goals conceded away
MV	pmnagh	percentage of matches with no goals conceded at home
MV	pmnaga	percentage of matches with no goals conceded away
LV	**Success**	
MV	Twh	total number of wins at home
MV	Twa	total number of wins away
MV	Tpts	total points at the end of season

Statistical analysis

The purpose of this study was to develop a team success model for a group of professional football teams that competed in the TSL during the 2016-2017 season. The process of modeling using Structural Equation Modeling (SEM) can be outlined in the following five steps: Model Specification, Model Identification, Estimation, Model Evaluation, and Model Modification. In general, models consist of both a measurement model and a structural model. The measurement model relates observed responses or indicators to LVs and sometimes to observed covariates (i.e., the CFA model). The structural model then specifies relations among LVs and regressions of LVs on observed variables or MVs. Figure 3 displays the LVs and MVs for the team success model.

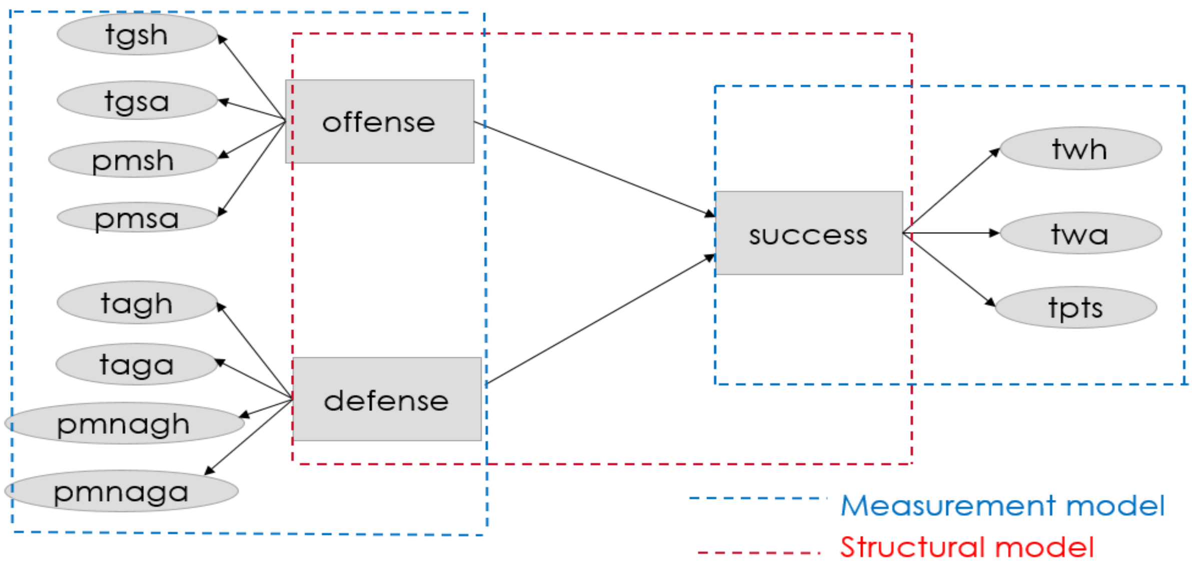

Figure 3. Structural (inner) and measurement (outer) parts of the team success model.

A specific SEM technique, which requires no distributional assumptions and is known as Partial Least Square-Structural Equation Modeling (PLS-SEM), was used here with a two-fold goal. First, it was necessary to identify a set of manifest variables (MVs), or indicators, to reflect the offensive and defensive latent variables (LVs) based on technical data obtained. Secondly, there was a need to examine and interpret the results of a PLS path model. In turn, this process required two steps. In the first step, the assessment of the measurement model, which is to verify that what was measured was what was intended to be measured, was carried out. In the second step, the assessment of the structural model, which was to draw conclusions about the relationships among the latent variables, was conducted.

Subsequently, the evaluation of a reflective measurement model is a three-fold process. First, the unidimensionality of the MVs was assessed. Second, the verification that the MVs were well explained by their associated LVs was conducted. Third, the assessment of the degree to which a given construct was different from other constructs was performed. The unidimensionality of the MVs, in turn, was assessed by the three main indices for unidimensionality, which are Cronbach's α coefficient, the Dillon-Goldstein's rho, and the first eigenvalue of the correlation matrix for the MVs. Cronbach's α coefficient is an average inter-variable correlation between the MVs of a reflective construct. The commonly acceptable value of α is .7 or higher (Chin, 2010). The Dillon-Goldstein's rho is another unidimensionality index that focuses on the variance of the sum of the variables in the block of MVs, with an acceptable value of rho being greater than .7. Additionally, to ensure unidimensionality, the first eigenvalue should be larger than 1, whereas the second eigenvalue should be smaller than 1 (Ravand & Baghaei, 2016).

PLS-SEM path modeling is implemented by using the R package *plspm*. I started out by estimating the LV scores to quantify the relationships in the model displayed in Figure 3. For each arrow, a numeric value representing the strength and direction of the relationship is obtained. The arguments to define the PLS path model are as follows: Data (the location and name of the data file), path matrix (definition of the structural model), blocks (a list defining the blocks of variables of the measurement model), scaling (a list defining the measurement scale of variables for non-metric data), and modes (a vector defining the measurement mode of each block). There are additional arguments for which I used the default values. The reader is referred to use *help(pslpm)* function to consult the technical document for the details of the other parameters.

	offense	*defense*	*success*
offense	0	0	0
defense	0	0	0
success	1	1	0

Figure 4. Structural (inner) model and measurement (outer) model for index of success.

The definition of the structural model represents the relationships among the LVs. Based on Figure 3, one can think of the structural model as a flowchart or network representing a causal process. Hence, it can be represented by a lower triangular Boolean type square matrix consisting of 0s and 1s, as depicted in Figure 4. For instance, the 1 in the cell (3,1) means that *offense* affects *success*. The 0s in the diagonal of the matrix mean that an LV cannot affect itself.

The definition of the measurement model is achieved by using a list containing vectors. Basically, the idea is to indicate the set of MVs that form each LV. In other words, I specify to the *plspm ()* function what variables of the data set are associated with what LVs. For example, the measurement model, which is coded in R as measuremodel<-list (1:4,5:8,9:11), communicates to the *plspm()* function that LV *offense* is associated with the first four columns of the data set, the LV *defense* is formed by columns from 5 to 9 of the data set, and the LV *success* is associated with columns from 9 to 11 in the data set. The PLS-SM path model is executed by running the segment of R code presented in Figure 5.

```
# Create the row vectors for path matrix (structural model)
offense<-c(0,0,0)
defense<-c(0,0,0)
success<-c(1,1,0)
# Define the structural (inner) model matrix
inmodel<-rbind (attack, defense, success)
colnames (inmodel)<-rownames (inmodel)
```

```
# Latent variables are measured in a reflective way
mode<-c("A","A","A")
# Specify the measurement (outer) model
measuremodel<-list(1:4,5:8,9:11)
plsmodel<-plspm(trlignum,inmodel,measuremodel,mode,boot.val=TRUE,br=200)
summary(plsmodel)
```

Figure 5. Segment of R code to execute the PLS-SEM path model.
Note: *mode defines the measurement model for each LV, boot.val indicates whether bootstrap validation must be performed, and br represents the number of bootstrap resamples.*

PLS-SEM path modeling follows a sequential procedure that can be divided in three phases, which are getting the weights to compute the LV scores, estimating the path coefficients for the structural model, and obtaining the loadings for the measurement model. The first phase consists of iteratively obtaining the weights to be used to get the scores of the LVs. The second phase has to do with estimating the path coefficients of the structural model. Finally, the third phase involves the computation of the loadings for the measurement model. As I aim to keep the mathematics at a minimum level in this paper, I refer the interested reader to Chin (2010) and Sanchez (2013), who have presented a detailed mathematical treatise of these phases by using various algorithmic schemes.

RESULTS

Having defined PLS-SEM path model and applied the function *plspm ()* to estimate the parameters, I present the interpretation of the results provided by the summary feature of the *plspm()* function in the next paragraphs. The findings are reported in the following order: (1) Summary statistics and normality measures, (2) Convergent validity of the MVs, (3) Discriminant validity of the MVs, (4) Measurement model evaluation by the reliability of the LVs, (5) Path coefficients for the SEM via bootstrapping, and (6) The predictability of the model by the R^2 for the endogenous LVs via bootstrapping and the overall model.

A summary of the descriptive statistics is displayed in Table 2. For all MVs, there were no values of skewness and kurtosis in excess of 1.96 (Field, Miles, and Field, 2012).

Table 2. Descriptive Statistics for MVs of the LVs: offense, defense, and success (n=18).

MVs	*M*	*SD*	*Skewness*	*Kurtosis*
Tgsh	25.833	8.410	0.761	-0.304
Tgsa	19.500	5.864	0.800	-0.397
Pmsh	72.222	11.196	1.274	1.264
Pmsa	65.111	13.429	0.293	-0.473
Tagh	20.278	6.488	0.352	-0.407
Taga	25.778	7.589	-0.088	-1.106
Pmnagh	33.389	14.880	0.480	-1.172
Pmnaga	25.611	16.067	0.568	-0.038
Twh	7.667	2.744	0.220	-0.377
Twa	5.611	2.682	0.469	-1.185
Tpts	47.278	14.708	0.598	-0.243

Note. tgsh=total number of goals scored at home, tgsa=total number of goals scored away, pmsh=percentage of matches with goals scored at home, pmsa=percentage of matches with goals scored away, tagh=total number of goals conceded at home, taga=total number of goals conceded away, pmnagh=percentage of matches with no goals conceded at home, pmnaga=percentage of matches with no goals conceded away, twh=total number of wins at home, twa=total number of wins away, and tpts=total points at the end of season

The quality of a measurement model can be assessed by measuring how much of the variance of the MVs of an LV is shared, which is referred to as the convergent validity, and is established by factor loadings higher than 0.7 (Ravand & Baghaei, 2016; Sanchez, 2013). The results, presented in Table 3, showed that only one of the MVs loaded lower than 0.7, while another one was very close to 0.7. These MVs were tgsa (0.665) for *offense*, and pmnagh (0.693) for *defense*. As these MVs had loadings that were very close to 0.7, I decided to keep them in the analysis. The loadings, weights, communalities, and redundancy for the MVs are given in Table 3.

Table 3. Loadings and Weight of MVs for LVs: offense, defense, and success.

MVs	*LV*	*loading*	*weight*	*Communality*	*redundancy*
Tgsh	offense	0.880	0.334	0.775	0.000
Tgsa	offense	0.665	0.235	0.442	0.000
Pmsh	offense	0.816	0.329	0.666	0.000
Pmsa	offense	0.844	0.333	0.712	0.000
Tagh	defense	0.794	0.323	0.631	0.000
Taga	defense	0.862	0.375	0.743	0.000
Pmnagh	defense	0.693	0.214	0.481	0.000
Pmnaga	defense	0.811	0.335	0.658	0.000
Twh	success	0.856	0.325	0.732	0.695
Twa	success	0.896	0.363	0.802	0.761
Tpts	success	0.997	0.394	0.993	0.942

Note. tgsh=total number of goals scored at home, tgsa=total number of goals scored away, pmsh=percentage of matches with goals scored at home, pmsa=percentage of matches with goals scored away, tagh=total number of goals conceded at home, taga=total number of goals conceded away, pmnagh=percentage of matches with no goals conceded at home, pmnaga=percentage of matches with no goals conceded away, twh=total number of wins at home, twa=total number of wins away, and tpts=total points at the end of season

In addition to convergent validity, discriminant validity is another tool used to assess how distinct a given LV is from another LV. The cross loadings of the MVs are given in Table 4. Because the loadings of the MVs associated with a given LV are higher than their loadings with any other LV, there are no traitor MVs present. All the MVs are loyal to their respective LVs. Therefore, the results displayed in Table 4 are in support of discriminant validity (Chin, 2010).

Table 4. Cross Loadings of MVs for each LV.

MVs	*LV*	*offense*	*defense*	*success*
Tgsh	offense	**0.880**	0.417	0.759
Tgsa	offense	**0.665**	0.233	0.442
Pmsh	offense	**0.816**	0.524	0.749
Pmsa	offense	**0.844**	0.423	0.758
Tagh	defense	0.308	**0.794**	0.649
Taga	defense	0.602	**0.862**	0.752
Pmnagh	defense	0.071	**0.693**	0.429
Pmnaga	defense	0.500	**0.811**	0.673
Twh	success	0.769	0.613	**0.856**
Twa	success	0.786	0.761	**0.896**
Tpts	success	0.861	0.836	**0.997**

A composite α value of 0.7 or higher provides evidence to support homogeneity of the MVs (Chin, 2010). The composite α values indicated good internal consistency for each LV, *offense* (α=0.82), *defense* (α= 0.87), and *success* (α= 0.90). Additionally, average variance extracted (AVE) was examined for measuring reliability of the LVs. The results showed that all AVEs were above 0.5, which indicated that more than 50% of the variance of the MVs of a given LV is shared. Hence, the AVE values greater than 0.5 indicated good convergent validity for each LV (Chin, 2010; Ravand and Baghaei, 2016). Furthermore, the measurement model can also be evaluated by the unidimensionality of the LVs. Unidimensionality of the measurement model can be examined by investigating if the first eigenvalues are great than 1 and the second eigenvalues are less than 1 for each LV (Ravand & Baghaei, 2016). Based on this criterion, Table 5 shows that first and second eigenvalues for all the LVs in our measurement model are reasonably within the acceptable range.

Table 5. Structural Model Correlations and Measurement Model Reliability Measures

AVE	**C. alpha**	**DG rho**	**1st Eigen**	**2nd Eigen**	**LVs**	**Mode**	**MVs**	***Offense***	***defense***	***success***
0.649	0.82	0.88	2.60	1.11	*offense*	A	4	1.000	-	-
0.628	0.80	0.87	2.53	1.03	*defense*	A	4	0.508	1.000	-
0.843	0.90	0.94	2.53	0.46	*success*	A	3	0.878	0.808	1.000

Note: AVE=Average Variance Extracted, C. alpha=Cronbach's alpha, DG rho= Dillon-Goldstein's rho, 1st Eigen=First Eigenvalue, 2nd Eigen=Second Eigenvalue, and Mode A=reflective constructs.

Having assessed the quality of the measurement model, the next I focused on evaluating the structural model. In order to inspect the results of each regression in the structural equations, the path coefficients presented in Table 6 were examined. The paths in the model from the LVs of *offense* (β= 0.630, p< .001) and *defense* (β= 0.489, p< .001) were statistically significantly and positively correlated with the LV of *success*. I used bootstrapping to create 200 samples of size n=18 resampled from the original data set to examine the variability and stability of the paths in the PLS-SEM path model. The significance of model parameter estimates was assessed at the significance level of α=.05. As displayed in Table 6, the results indicated statistically significant confidence intervals for the paths in the model.

Table 6. Structural Model Path Coefficients

Paths for LVs	β	*SE*	*t*	Mean Boot	*SE*	95% *CI*	
offense->success	0.630	0.068	9.26***	0.615	0.114	0.446	0.742
defense->success	0.489	0.068	7.18***	0.495	0.086	0.344	0.685

Note: ***p<.001.

Besides the results of the regression equations, the quality of the structural model is evaluated by examining three indices or quality metrics. These indices are the coefficient of determination, R^2, the redundancy index, and the Goodness-of-Fit (GoF) index. The coefficient of determination, R^2, is used to assess the quality of the structural model. As displayed in Table 7, R^2 for *success*, as predicted by the two LVs of *offense* and *defense*, was 0.948. According to the effect size index established by Cohen (1988), the endogenous LV of *success* have a large effect size.

Table 7. Original and Bootstrap R^2 Results for Structural Model Summary

LVs	Type	R^2	Mean Boot	*SE*	95% CI		Block Comm
offense	Exogenous	0.000					0.649
defense	Exogenous	0.000					0.628
success	Endogenous	0.948	0.435	0.034	0.367	0.501	0.843

Note. Block Comm=LV Communality.

Redundancy is a measure of the amount of variance in an endogenous LV explained by its exogenous LVs. In other words, it reflects the ability of a set of exogenous LVs to explain variation in an endogenous LV. Hence, a high redundancy value indicates a high predictive ability of the model. The redundancy for the overall model was found to be 0.799. In the model presented here almost 80% of the variation in the LV *success* could be explained by the LVs *offense* and *defense*. In other words, the exogenous LVs of *offense* and *defense* in the model presented here predicted 79.9% of the variability of the LV *success* and its MVs.

As there are no inferential tests for the Goodness-of-Fit (GoF) in PSL, it is a pseudo goodness of fit measure that accounts for the model quality for the measurement and structural models, and it is calculated as the geometric mean of the average communality and the average R^2 value (Sanchez, 2013). The GoF R^2 for the overall model was found to be 0.811. Chin (1998) classified R^2 values of less than 0.19 as weak, greater than 0.33 as moderate, and greater than 0.67 as substantial, while Sanchez (2013) considered the R^2 values between 0.3 and 0.6 to be moderate. Hence, the overall GoF R^2 value of 0.811 for the model presented here is substantially high.

DISCUSSION

In this study, by using the context of the Turkish Super League (TSL) in the 2016-2017 season, I aimed to demonstrate the basics of how to implement a PLS-SEM path analysis technique. To my knowledge, this is one of the only two studies that implemented PSL-SEM techniques to investigate and identify indicators for team success in the context of professional football. I considered several MVs in combination and their respective LVs to assess team success in the TSL competition. The MVs were grouped in Offense Related Variables (ORVs), Defense Related Variables (DRVs) and Success Related Variables (SRVs). However, I did not make any differentiation in terms of the variability of offensive schemes or the compactness of defensive structures used by the teams.

The exogenous LVs of *offense* (β= 0.630, p< .001) and *defense* (β= 0.489, p< .001) both had statistically significant positive effects on the endogenous LV of *success*. Hence, the defensive ability of a team seemed just as important as the offensive ability for team success in football. This particular result is well aligned with the outcome of various studies conducted by other researchers (Dufour, et al., 2017). Furthermore, Hughes & Churchill (2004) stated that in their study it appeared that defensive ability of teams to control the opposing team's movements had a significant effect on team success.

Based on the PLS-SEM path model presented here, one can obtain a predicted ranking of the teams in the TSL for the 2016-2017 season. This is achieved in PLS-SEM path modeling by calculating the scores for the endogenous LV *success* in the model. Subsequently, a ranking of the teams based on their model predicted success scores is constructed. Comparing this model predicted rank order to the actual ranking table at the end of the 2016-2017 season, one can draw conclusions as to how well the model predicts. When the teams were ranked based on their model predicted success scores, actual end of season rankings of 10 out of 18 teams were predicted accurately by the PLS model. Out of the 8 teams, 5 teams had higher model predicted rankings than that of their actual rankings. This may lead to a conclusion that when the model did not predict accurately, the model showed a small tendency to slightly higher rank the teams in comparison with their actual rankings. Table 8 displays the scores for the LVs, as well as the model predicted and actual rankings.

Table 8. Scores for the LVs in the model and ranking of the teams.

Team	Model Rank	offense	defense	success	Actual Rank	Δ Rank
Beşiktaş	1	2.559	1.342	2.089	1	0
Istanbul Başakşehir	2	1.206	1.661	1.716	2	0
Galatasaray	3	1.534	0.210	1.344	4	-1
Fenerbahçe	4	1.001	1.060	1.118	3	1
Antalyaspor	5	0.342	0.495	0.794	5	0
Trabzonspor	6	-0.765	1.235	0.234	6	0
Akhisar Belediye	7	0.013	0.956	0.097	7	0
Gençlerbirligi	8	-0.852	0.980	-0.220	8	0
Kasımpaşa	9	-0.017	-0.375	-0.303	9	0
Kardemir Karabük	10	-0.956	-0.227	-0.339	11	-1
Alanyaspor	11	0.465	-1.270	-0.387	12	-1
Konyaspor	12	-0.470	-0.249	-0.407	10	2
Bursaspor	13	-0.686	-0.836	-0.564	15	-2
Kayserispor	14	-0.125	-1.257	-0.686	14	0
Rizespor	15	-0.347	-0.969	-0.742	16	-1
Osmanlıspor	16	-0.560	-0.103	-0.772	13	3
Gaziantepspor	17	-1.491	-1.229	-1.403	17	0
Adanaspor	18	-0.851	-1.424	-1.570	18	0

Note. Δ Rank=Model Rank – Actual Rank.

Based on Table 8, there were only three teams for which the difference between the predicted and the actual ranking was 2 or 3, in either direction. It is noted that the model predicted the top and the bottom two team rankings correctly. Furthermore, the model also quite accurately predicted the rankings of the top 10 teams.

CONCLUSIONS

In this study, the focus was on the TSL during the 2016-2017 season, and the results may be restricted to this level of competition of this era. Hence, the results of this study may not be generalizable to other types of competitions, such as the EPL, Spanish La Liga, Italian Serie A, Champions League, or especially to FIFA World Cup, as it is played in a much shorter format compared to a typical league format. Needless to say, the role of chance, especially in competitions with a short format, as a determinant of team success is considerably magnified.

As I implemented the theory of team success in football proposed by Wade (1996), my main focus was on the two of the three fundamental phases of play, attack or offense and defense. The third phase of play, the preparation phase, which is sometimes also referred to as midfield play, is based on passing sequences and ball possession, which are, in turn, can be molded by a specific coaching and/or tactical philosophy and style of play, was not a part of the model examined here. In the future, there is a need for a research study to enhance the model presented here. This can be achieved by conducting an investigation into identifying a set of indicators for the preparation phase of play and incorporating it into the model presented here.

PRACTICAL APPLICATION

Even though goals scored and conceded by teams at home and away are variables that could be easily measured to determine team success, the goal scoring, in general, is a relatively low frequency event in football. Hence, as suggested by Lago-Peñas & Martin (2007), the narrow range of values and variance may prevent one from conducting robust variance-covariance based multivariate analyses without using scaling and/or transformations, especially with respect to other variables. Furthermore, other team success indicators, such as the number and nature of shots on goal and their outcomes, need to be included in the model to identify patterns of behaviors related to successful team performance.

Although there are other external factors that may have an impact on team success, such as opposition effects, match officials and venue, these factors were beyond the scope of this research study. For instance, as for the opposition effects, the rankings of the teams competing against each other were not included among the variables in this study. Future research studies can take the team rankings into account when refining the index of success model presented here in order to develop an improved model.

In closing, from a philosophical stand point, the "truth", or full reality, if one prefers, in biological and social sciences has essentially infinite dimension(s). Hence, the full reality cannot be revealed with only finite samples of data and/or a model constructed based on these data (Burnham & Anderson, 2002). It is also worth remembering that the basic rule of statistical modeling was best summarized by Box (1976), who surmised that all models are wrong, but some are useful.

REFERENCES

Armatas, V., Yiannakos, A. & Sileloglou, P. (2007). Relationship between time and goal scoring in soccer games: Analysis of three World Cups. *International Journal of Performance Analysis in Sport*, 7(2), 48–58.

Box, G.E.P. (1976). Science and statistics. *Journal of the American Statistical Association, 71*(356), 791–799.

Burham, K.P., Anderson, D.R. (2002). Model selection and multi-model inference. Springer-Verlag, New York.

Carling, C., Le Gall, F., McCall, A., Nédélec, M. & Dupont, G. (2015). Squad management, injury and match performance in a professional soccer team over a championship-winning season, *European Journal of Sport Science*, *15*(7), 573–582.

Castellano, J., Casamichana, D. & Lago, C. (2012). The use of match statistics that discriminate between successful and unsuccessful soccer teams. *Journal of Human Kinetics*, *31*, 139–147.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed), Modern methods for business research, 295–358. Mahwah, NJ: Lawrence Erlbaum.

Chin, W. W. (2010). How to write up and report PLS analyses. In Vinzi, V. E., Chin, W. W., Henseler, J., and Wang, H. (Eds), Handbook of partial least squares: Concepts, methods and applications in marketing and related fields, 655–690. Berlin: Springer.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Crawley, M.J. (2007). The R Book. John Wiley & Sons, Ltd., Chichester.

Dufour, M., Phillips, J. & Ernwein, V. (2017). What makes the difference? Analysis of the 2014 World Cup. *Journal of Human Sport and Exercise, 12*(3), 616–629. doi:https://doi.org/10.14198/jhse.2017.123.06

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 1-15). Thousand Oaks, CA, US: Sage Publications, Inc.

Hughes, M.D., Bartlett, R.M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences, 20*(10), 739–754.

Hughes, M.D., Churchill, S. (2004). Attacking profiles of successful and unsuccessful teams in Copa America 2001. *Journal of Sport Sciences, 22*(6), 505.

Hughes, M.D., Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, *23*(5), 509–514.

James, N., Jones, P.D. & Mellalieu, S.D. (2004). Possession as a performance indicator in soccer as a function of successful and unsuccessful teams. *Journal of Sport Sciences, 22*(6), 507–508.

Fernandez-Navarro, J., Fraduab, L., Zubillagac, A., Forda, P.R. & McRoberta, A.P. (2016). *Journal of Sports Sciences*, *34*(24), 2195–2204.

Field, A., Miles, J. & Field, Z. (2012). Discovering statistics using R. London: Sage.

Jones, P.D., James, N., and Mellalieu, S.D. (2004). Possession as a performance indicator in soccer. *International Journal of Performance Analysis in Sport, 4*, 98–102.

Kempe, M., Vogelbein, M., Memmert, D. & Nopp, S. (2014). Possession vs. direct play: Evaluating tactical behavior in elite soccer. *International Journal of Sports Science, 4*(6A), 35–41.

Lanham, N. (2005). The goal complete: The winning difference. Science and football V. London: Routledge.

Lago-Peñas, C., Dellal, A. (2010). Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *Journal of Human Kinetics, 25*, 93–100.

Lago-Peñas, C., Martin, R. (2007). Determinants of possession of the ball in soccer. *Journal of Sports Sciences, 25*(9), 969–974.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ravand, H., Baghaei, P. (2016). Partial least structural equation modeling with R. *Practical Assessment, Research & Evaluation, 21*(11), 1–16.

Saito, K., Yoshimura, M. & Ogiwara, T. (2013). Pass appearance time and pass attempts by teams qualifying for the second stage of FIFA World Cup 2010 in South Africa. *Football Science, 10*, 65–69.

Sanchez, G. (2013) PLS Path Modeling with R. Trowchez Editions. Berkeley, 2013. http://www.gastonsanchez.com/PLS Path Modeling with R.pdf

Sanchez, G., Trinchera, L. & Russolillo, G. (2017). plspm: Tools for Partial Least Squares Path Modeling (PLS-PM). R package version 0.4.9. https://CRAN.R-project.org/package=plspm

Scoulding, A., James, N. & Taylor, A. (2004). Passing in the soccer world cup 2002. *International Journal of Performance Analysis in Sport, 4*(2), 36–41.

Szwarc, A., (2007). Efficacy of successful and unsuccessful soccer teams taking part in finals of champions league. *Research Yearbook, 13*(2), 221–225.

Tenga, A., Holme, I., Ronglan, L.T. & Bahr, R. (2010). Effect of playing tactics on goal scoring in Norwegian professional soccer. *Journal of Sports Sciences, 28*(3), 237–244.

Tenga, A., Sigmundstad, E. (2011). Characteristics of goal-scoring possessions in open play: Comparing the top, in-between and bottom teams from professional soccer league. *International Journal of Performance Analysis in Sport, 11*(3), 545–552.

Wade, A. (1996). Principles of team play. Spring City: Reedswain.

CITATION OF THIS ARTICLE

Türegün, M. (2019). Partial Least Squares-Structural Equation Modeling (PLS-SEM) Analysis of Team Success Using R, International Journal of Sport, Exercise & Training Sciences - IJSETS, 5(4), 201–213. DOI: 10.18826/useeabd.628653