# *Competitive Balance in Football Leagues: Domestic vs International*

Aman Sinha, Rohit Agarwal, Utkarsh Bairolia, Vishal Sourav

MathSport Asia | Conference 2018

# Content

- Problem Description
- Targets
- Factors and Parameters
- Challenges
- Correlation
- Model
- Inferences
- Future Prospects
- References

# Problem Description

The overall success of a football event involves both sport and economic success. One of the major factor affecting the success of football event is competitiveness among teams.

*“Appropriate measurement of competitive balance is a cornerstone of the economic analysis of professional sports leagues.” (Owen P. and King N. 2013)*

Higher levels of competitive balance will lead to higher levels of revenue. Quantification/ Measure of competitive balance (competitiveness) can be useful in order to compare and forecast the behaviour of leagues.

# Targets

1. Our main aim is to capture the seasonal fluctuation in popularity of leagues by comparing it using various competitiveness parameters.
2. A model that shows the relative competitiveness among domestic football leagues and international league.

**Domestic Leagues**: SPL, EPL, Bundesliga, Ligue 1, Serie A

**International Leagues**: UEFA Champions League

# Factors and Parameters

- CH5 (originally known as C5)
- CH5ICB (originally known as C5ICB)
- H
- HICB
- Upsets
- Relative Entropy
- Win% Range
- CL5
- CL5ICB
- Standard Ratio
- Standard Win%

# Factors and Parameters

- **CH5**- It is the ratio of total points occupied by top 5 clubs and total points of league. It represents the dominance of top 5 clubs in league. Higher the value of CH5, less the competitive balance.

CH5 Ratio=(Total points occupied by top five clubs) / (Total number of points occupied by all clubs)

- **CH5ICB**- It is defined to balance the change in CH5 due to change in number of clubs in league.

CH5 Index of Competitive Balance = [ CH5 / (5 / N) ]* 100

# Factors and Parameters

- **H**- It represents the inequality between clubs. It reflects the degree of competitiveness among all clubs. Higher the value of H index, less the competitive balance.

  $H=\sum s^2$ where s= club's share of points.

- **HICB**- It is defined to balance the change in H index due to change in number of clubs in league.

  $HICB= [H / (1 / N)] * 100$

  where N=Number of clubs in league

# Factors and Parameters

- **Upsets**-If a lower order club wins against higher order club, it is termed as upsets. Higher the upsets, more the competitiveness.

- **Relative Entropy**- It is defined to measure the imbalance between clubs.

$$\text{Relative Entropy} = -\sum P_i \log_2 P_i \,/\, \log_2 N$$

where $P_i$ = proportion of league victories of ith club and N is the number of clubs in the league.

# Factors and Parameters

- **Win% Range**- It is defined as the difference between highest winning percentage and lowest winning percentage. Higher the win% range, less the competitive balance.

Win% = (Total points won by a club) / (Total number of games played)* 3

Win% Range = Difference between highest and lowest win%

# Factors and Parameters

- **CL5**- It is defined as a measure to calculate the accountancy of bottom 5 clubs in league. Higher the CL5, better the competitive balance.

CL5 = (Total points occupied by bottom 5 clubs)/(Total number of points occupied by all clubs)

- **CL5ICB**- It is defined to balance the change in CL5 due to change in the number of clubs.

$$CL5ICB=[\ CL5\ /\ (5\ /\ N)\ ]*\ 100$$

where N = Number of clubs in the league

# Factors and Parameters

- **Standard Ratio**- It is the ratio of actual standard ratio to an idealized standard ratio. Higher the ratio, less the competitive balance.

- **Standard Win%**- It is the standard deviation of win% in a season. Higher the standard win%, less the imbalance between the clubs and hence less the competitive balance.

# Data & Challenges

- The analysed dataset contains the scorecard of different clubs in a particular league.
- The initial set of feature is composed of attributes like Date, HomeTeam, AwayTeam, Full Time Home Team Goals, Full Time Away Team Goals, Full Time Result, etc.
- *Each year is considered as an instance for a league*.
- Total points of a club is calculated by awarding 3 points for a win, 1 point to both the club for a tie and 0 points for defeat.

# Data & Challenges

- Each year of league is equivalent to an instance so even if we consider data from 1995-2018, there will be only 23 data sets which is not sufficient enough for generalisation of the model.

# Correlation

| | CH5 | CH5ICB | CL5 | CL5ICB | HI | HICB | Relative Entropy | Revenue | Std Ratio | Std Win | Upsets | Win Percentage Range |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CH5** | 1 | 0.45 | 0 | -0.40714 | 0.664286 | 0.735714 | -0.56429 | 0.042857 | **0.478571** | 0.585714 | 0.028571 | 0.428571 |
| **CH5ICB** | 0.45 | 1 | -0.61429 | -0.50714 | -0.10714 | 0.542857 | 0.207143 | 0.278571 | **0.671429** | 0.514286 | -0.3 | 0.307143 |
| **CL5** | 0 | -0.61429 | 1 | 0.721429 | 0.442857 | -0.22857 | -0.40714 | -0.49286 | **-0.37857** | -0.15714 | 0.35 | -0.07143 |
| **CL5ICB** | -0.40714 | -0.50714 | 0.721429 | 1 | -0.08571 | -0.45 | 0.121429 | -0.42857 | **-0.35714** | -0.32143 | 0.271429 | -0.18571 |
| **HI** | 0.664286 | -0.10714 | 0.442857 | -0.08571 | 1 | 0.607143 | -0.86429 | -0.23571 | **0.214286** | 0.471429 | 0.271429 | 0.235714 |
| **HICB** | 0.735714 | 0.542857 | -0.22857 | -0.45 | 0.607143 | 1 | -0.44286 | 0.1 | **0.742857** | 0.835714 | -0.09286 | 0.507143 |
| **Relative Entropy** | -0.56429 | 0.207143 | -0.40714 | 0.121429 | -0.86429 | -0.44286 | 1 | 0.192857 | **-0.17857** | -0.42857 | -0.2 | -0.20714 |
| **Revenue** | 0.042857 | 0.278571 | -0.49286 | -0.42857 | -0.23571 | 0.1 | 0.192857 | 1 | **0.128571** | 0.085714 | -0.09286 | 0.021429 |
| **Std Ratio** | 0.478571 | 0.671429 | -0.37857 | -0.35714 | 0.214286 | 0.742857 | -0.17857 | 0.128571 | **1** | 0.885714 | -0.13571 | 0.535714 |
| **Std Win** | 0.585714 | 0.514286 | -0.15714 | -0.32143 | 0.471429 | 0.835714 | -0.42857 | 0.085714 | **0.885714** | 1 | -0.13571 | 0.614286 |
| **Upsets** | 0.028571 | -0.3 | 0.35 | 0.271429 | 0.271429 | -0.09286 | -0.2 | -0.09286 | **-0.13571** | -0.13571 | 1 | -0.03571 |
| **Win Percentage Range** | 0.428571 | 0.307143 | -0.07143 | -0.18571 | 0.235714 | 0.507143 | -0.20714 | 0.021429 | **0.535714** | 0.614286 | -0.03571 | 1 |

# Model Approach & Selection

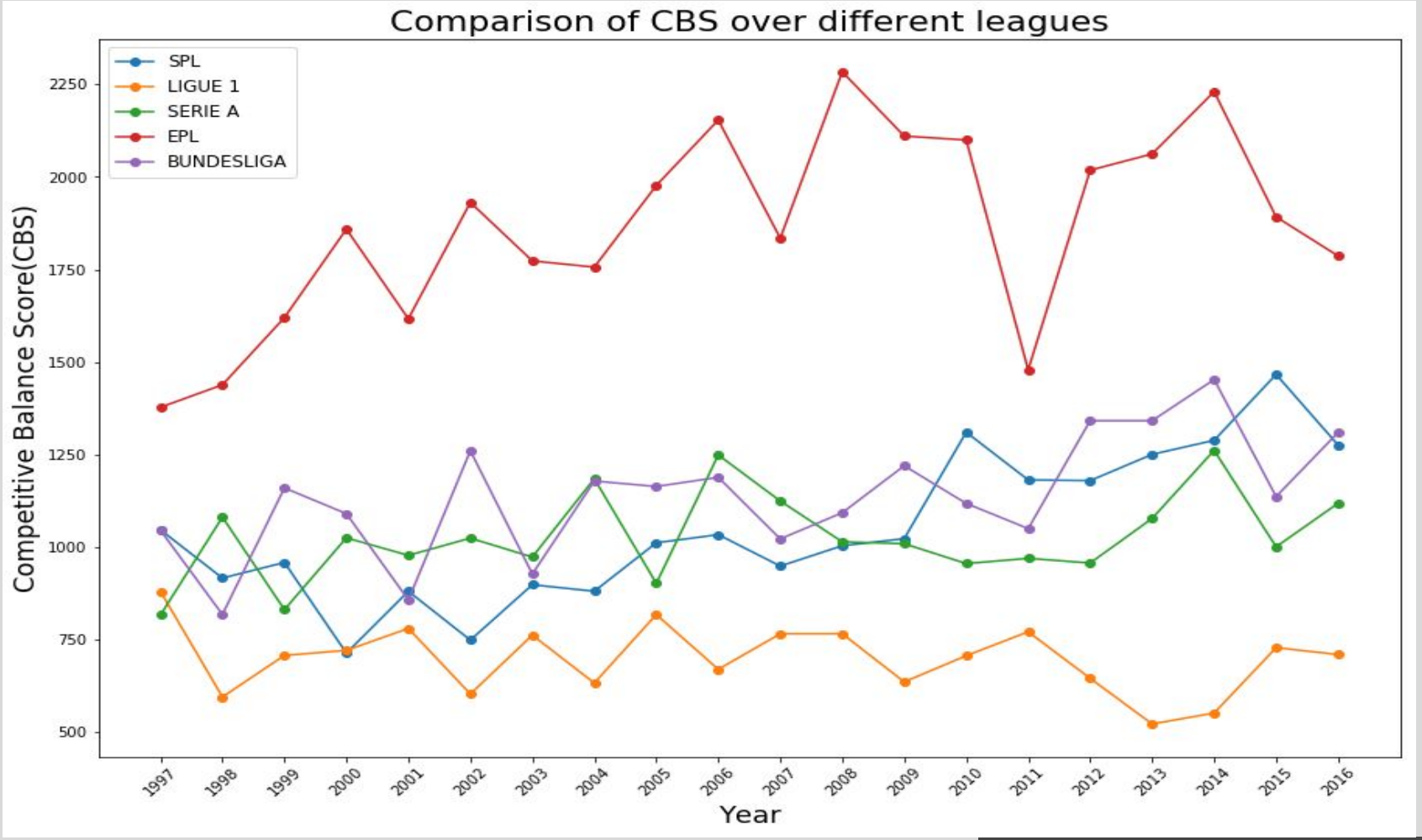
- Using the correlation matrix we selected the parameters which were highly correlated with others.
- Linear regression was applied on the datasets using the selected parameters.
- We selected different models on the basis of amount of significance and R-squared value.
- Considering ***effective revenue*** as the objective function, we train the model using above parameters and then we compare the findings with revenue.

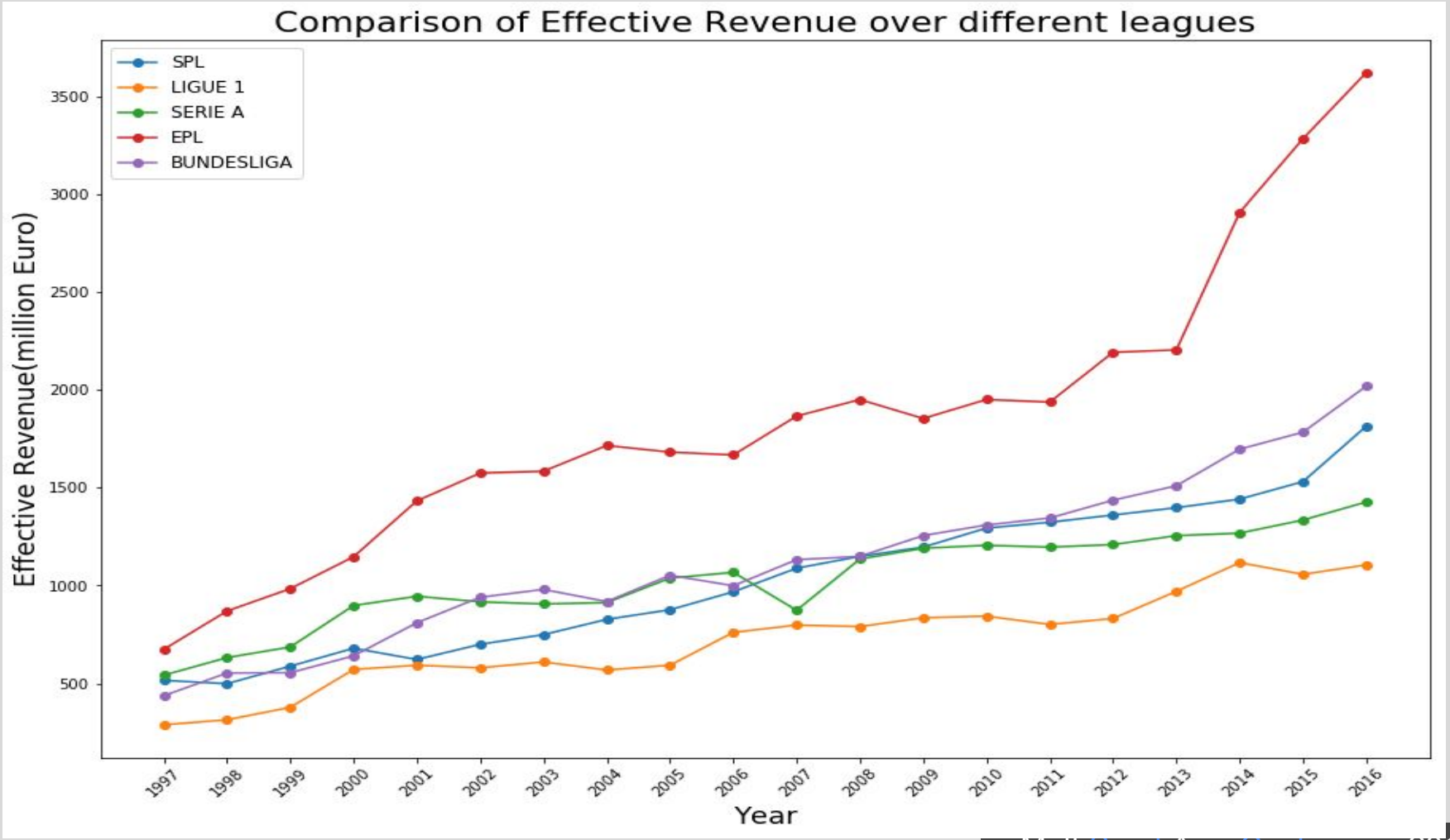
# Various Models

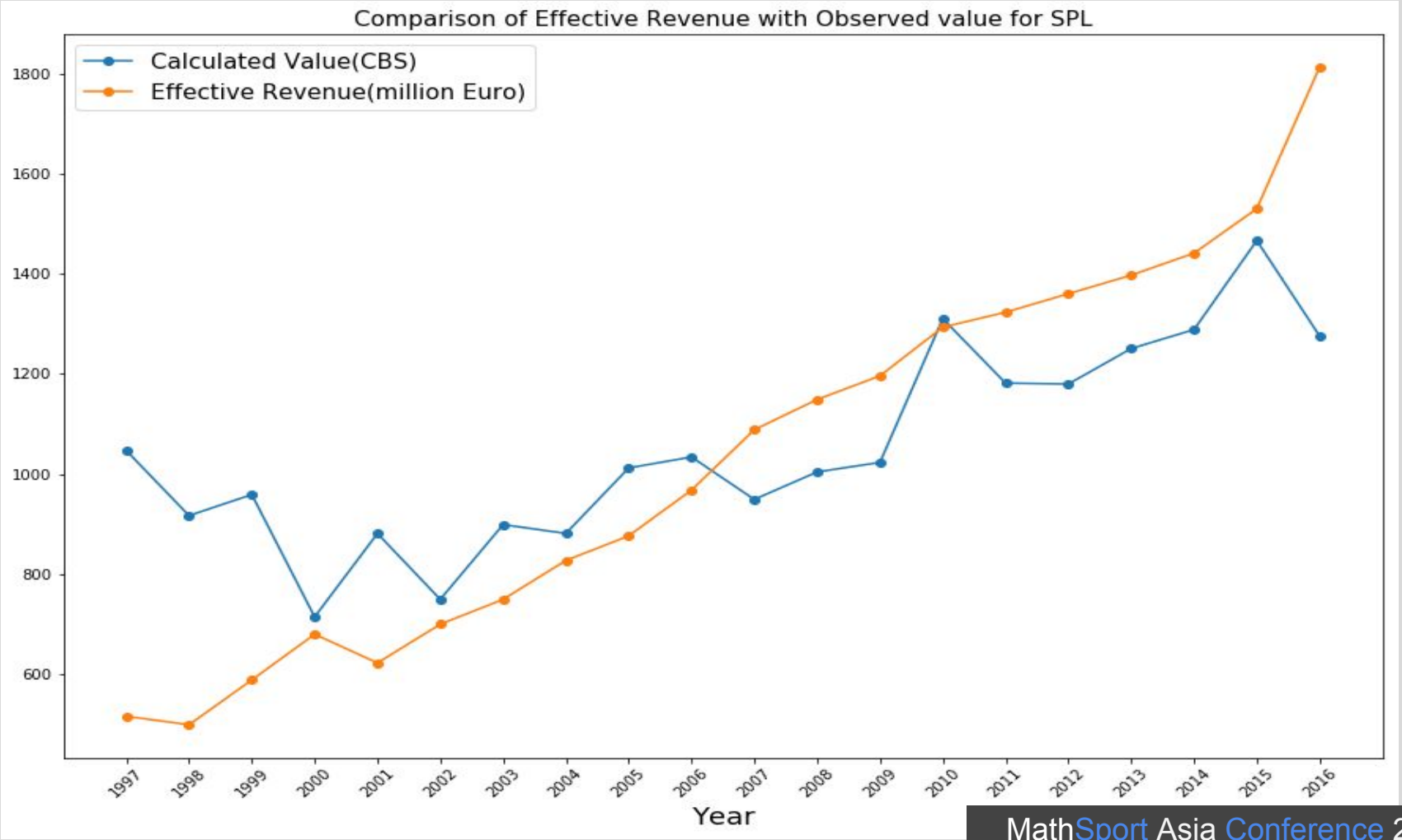
| Models | Parameters | P value | Multiple R-squared |
|---|---|---|---|
| Model 1 | Std Ratio | 6.58E-14 | 0.951 |
| Model 2 | HICB | 7.28E-11 | 0.897 |
| Model 3 | Std Ratio | 3.45E-05 | 0.961 |
| | HICB | 0.0393 | |
| Model 4 | Std Ratio | 3.42E-03 | 0.9622 |
| | HICB | 7.22E-02 | |
| | Win% Range | 6.09E-01 | |

CBS= (Coefficient)x(Standard Ratio)

| Leagues | Coefficient | Multiple R-squared |
|---|---|---|
| SPL | 921.418 | 0.951 |
| LIGUE 1 | 573.624 | 0.863 |
| SERIE A | 848.089 | 0.956 |
| EPL | 1513.316 | 0.893 |
| BUNDESLIGA | 1034.825 | 0.917 |
| UEFA Champions League | 554.168 | 0.807 |

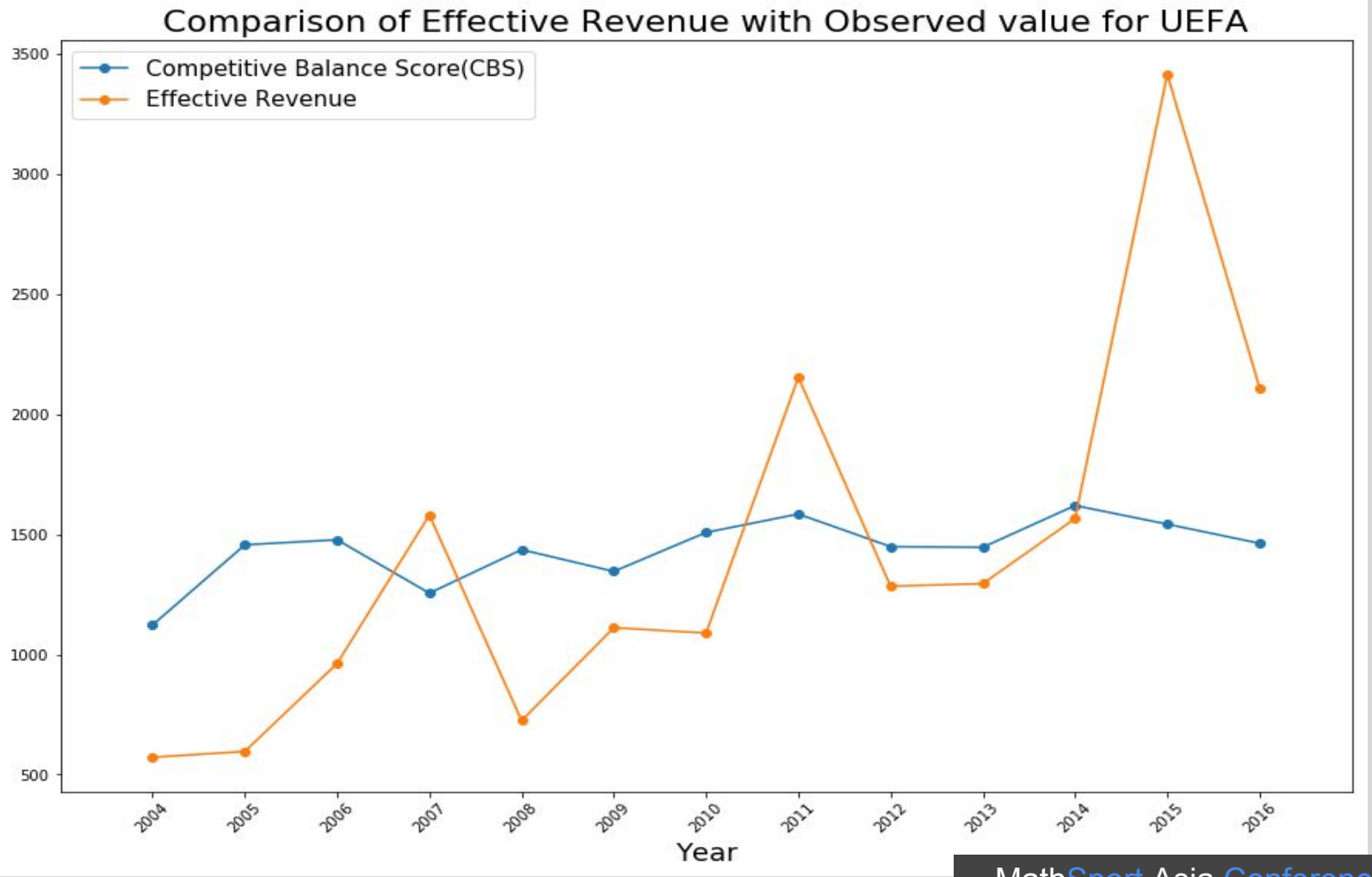

Comparison of CBS over different leagues
SPL
LIGUE 1
SERIE A
EPL
BUNDESLIGA
Competitive Balance Score(CBS)
2250
2000
1750
1500
1250
1000
750
500
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year


Comparison of Effective Revenue over different leagues
SPL
LIGUE 1
SERIE A
EPL
BUNDESLIGA
Effective Revenue(million Euro)
500
1000
1500
2000
2500
3000
3500
Year
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016


Comparison of Effective Revenue with Observed value for SPL
Calculated Value(CBS)
Effective Revenue(million Euro)
1800
1600
1400
1200
1000
800
600
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year

# International Leagues

1. The same model is applied for international league (UEFA Champions League) and the observed value calculated using model is plotted against the effective revenue.
2. It is observed from the plot that revenue is not the only factor accountable for competitive balance of a league.


Comparison of Effective Revenue with Observed value for UEFA
Competitive Balance Score(CBS)
Effective Revenue
3500
3000
2500
2000
1500
1000
500
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
Year

# Inferences

1. From our model, we can infer which league is more likely to generate the highest revenue among the leagues.
2. On the basis of revenue, we can't tell the competitive balance of a league for a particular season.

# Future Prospects

1. The data can be increased by considering month/ quarter as an instance in order to better generalise the model.
2. Our proposed model can be used to define a cost function to quantify competitiveness value.
3. Revenue along with some other physical parameters need to be considered in order to study the behaviour of competitiveness among the leagues.

# References


- Michie J. and Oughton C. 2004 *Competitive Balance in Football: Trends and Effect*
- Goossens K. 2006 *Competitive Balance in European Football: Comparison by Adapting Measures: National Measure of Seasonal Imbalance and Top 3*


**Stats and Dataset:**


- https://www.statista.com/
- http://www.football-data.co.uk/

# Thank You!