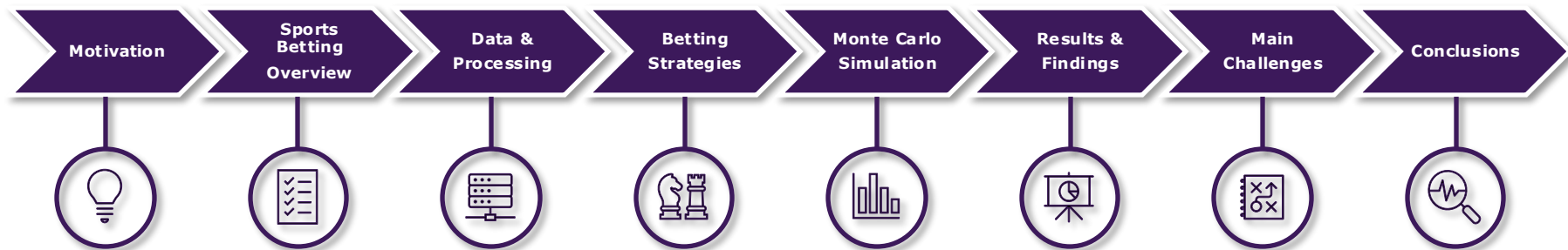# Betting in the Premier League

Advanced Python Programming for Economics, Management and Finance

Anna Katharina Schnabel – 3215884
Andrei Arin Pacuretu – 3200338
Ferran Garcia Rovira – 3329494
David Ponti Pascual – 3329041
Florian Gaszner – 3221668

Outline

Motivation

Sports Betting Overview

Data & Processing

Betting Strategies

Monte Carlo Simulation

Results & Findings

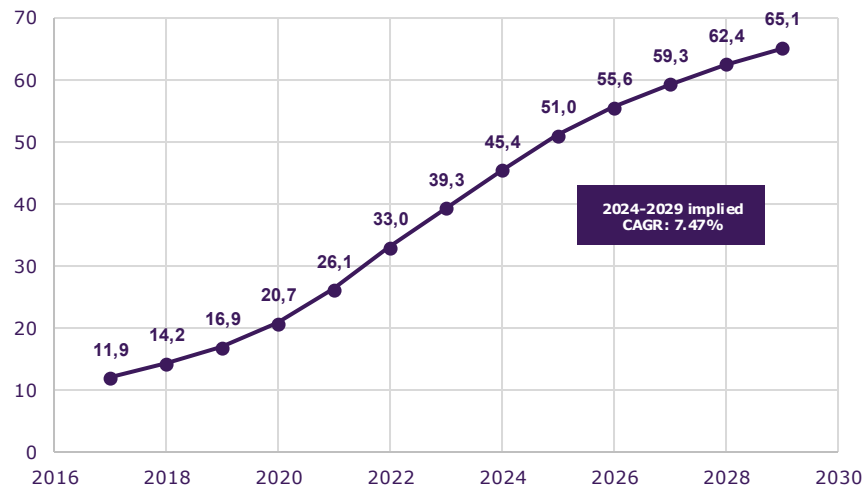Main Challenges

Conclusions

Motivation

Motivation

Why this topic is relevant

- The rise of sports betting and debates on market efficiency

- *"Do betting odds fully reflect available information, or are there exploitable inefficiencies?"*

- **Objective**: Investigating betting inefficiencies in Premier League odds (2005-2024) using data analysis and predictive models

Sports Betting Industry Expectations

Gaming Industry Revenue (in billion USD)

2024-2029 implied CAGR: 7.47%

Data points: 11,9 | 14,2 | 16,9 | 20,7 | 26,1 | 33,0 | 39,3 | 45,4 | 51,0 | 55,6 | 59,3 | 62,4 | 65,1

Premier League

Sports Betting as a Market

Sports Betting as a Market

Parallels

- **Key Concept**: Betting markets are similar to financial markets; odds are akin to prices, reflecting probabilities

- **Efficient Market Hypothesis**: Odds should represent all available information

- **Biases in the Industry**:
  - **Favorite-Longshot Bias:** Overvaluation of underdogs, undervaluation of favorites
  - **Home Bias:** Bettors overestimate the home advantage
  - **Sentiment Bias:** Odds skewed toward popular teams

- **COVID-19 Impact**: Margins increased due to unpredictable conditions

Largest Sports Betting Markets

High
Medium
Low

Data Overview and Processing

Raw Dataset

Original Variables

- **Date & Time**
- **Home & Away Teams**
- **FTHG & FTAG, HTHG & HTAG:** # of goals scored by home and away teams in Fulltime and at Halftime
- **FTR, HTR:** Fulltime and Halftime results (Home, Away, Draw)
- **Other relevant game statistics:** Shots, Shots on Target, Fouls, Corners, Yellow Cards, Red Cards – all for both Home and Away
- **Referee**
- **Odds by 5 bookmakers for the results**

Example of Dataset

Date	Home Team	Away Team	FTHG	FTAG	FTR
11/08/2023	Burnley	Man City	0	3	A
12/08/2023	Everton	Fulham	0	1	A

HS	AS	HST	AST	HC	AC
6	17	1	8	6	5
19	9	9	2	10	4

B365H	B365D	B365A	BWH	BWD	BWA
8.00	5.50	1.33	8.75	5.25	1.34
2.20	3.40	3.30	2.20	3.40	3.25

Data Overview

Our Dataset & Code Explanation

- **Dataset**: Matches from 2005-2024 (Premier League)
  - Odds data sourced from 5 bookmakers
  - Variables: home/away odds, attendance, outcomes

- **Code explanation:**
  - Read the CSV file into a pandas DataFrame (df)
  - Then we save the DataFrame as a global variable
  - Afterwards, we list the created variables

Code

```python
for file_name in sorted(os.listdir(data_dir)):

    if file_name.endswith(".csv") and file_name.startswith("Prem"):

        season = file_name.split(" ")[1].split(".")[0].replace("-", "")

        key = f"pl_{season}"

        file_path = os.path.join(data_dir, file_name)

        try:

            df = pd.read_csv(file_path, encoding='utf-8')

        except UnicodeDecodeError:

            df = pd.read_csv(file_path, encoding='latin-1')

        globals()[key] = df

        print(f"Archive {file_name} uploaded as {key}")


created_vars = [var for var in globals() if var.startswith("pl_")]

print("Created variables:", created_vars)
```

Data Cleaning Process

What Was Adjusted

- **Data Limitations:** Missing odds data for older matches required interpolation

- **Code Explanation:**
  - We drop columns not central to our analysis, such as Referee and Half Time Result
  - We filled in missing odds data with Bet365 odds, since this betting platform has more complete data

- For strategies betting on attendance of games, years without attendants (Covid-19) were excluded

Code

```python
if original_var in globals():

    Df = globals()[original_var]

    globals()[filtered_var] = df.drop(columns=columns, errors='ignore')
...

pl_2324_f["IWH"] = pl_2324_f["IWH"].fillna(pl_2324_f["B365H"])

pl_2324_f["IWD"] = pl_2324_f["IWD"].fillna(pl_2324_f["B365D"])

pl_2324_f["IWA"] = pl_2324_f["IWA"].fillna(pl_2324_f["B365A"])

pl_2324_f["BWH"] = pl_2324_f["BWH"].fillna(pl_2324_f["B365H"])

pl_2324_f["BWD"] = pl_2324_f["BWD"].fillna(pl_2324_f["B365D"])

pl_2324_f["BWA"] = pl_2324_f["BWA"].fillna(pl_2324_f["B365A"])
```

Betting Strategies

Betting Strategies General Setup

Setup

- **Code Explanation:**
  - Define function to calculate average odds based on the individual odds shown by the main 5 betting platforms
  - We also define a function for implied probabilities: We show implied odds for the home team winning but also calculate it for away and draw. Inverse of home odds divided by the sum of the inverse of home, away and draw odds.
  - Iterate average and implied odds on each data frame

- **Note:** We will in the following create several dummy variables
  1. **Betting dummies:** 1 if we bet on something
 2. **Bet result dummies:** 1 if we won the bet

Code

```python
def calculate_avg_odds(df, odds_columns, avg_column_name):

df[avg_column_name] = 0.2 * df[odds_columns].sum(axis=1)


def calculate_implied_prob(df, avg_home_odds, avg_draw_odds, avg_away_odds,

home_prob_col, draw_prob_col, away_prob_col):

        df[home_prob_col] = (1 / df[avg_home_odds]) / ((1 / df[avg_home_odds]) +

        (1 / df[avg_draw_odds]) + (1 / df[avg_away_odds]))
...
calculate_avg_odds(df, home_odds_columns, "Avg Home Odds")

calculate_avg_odds(df, draw_odds_columns, "Avg Draw Odds")

calculate_avg_odds(df, away_odds_columns, "Avg Away Odds")
...
calculate_implied_prob(df, "Avg Home Odds", "Avg Draw Odds", "Avg Away Odds",

"Implied Home Prob", "Implied Draw Prob", "Implied Away Prob")
```
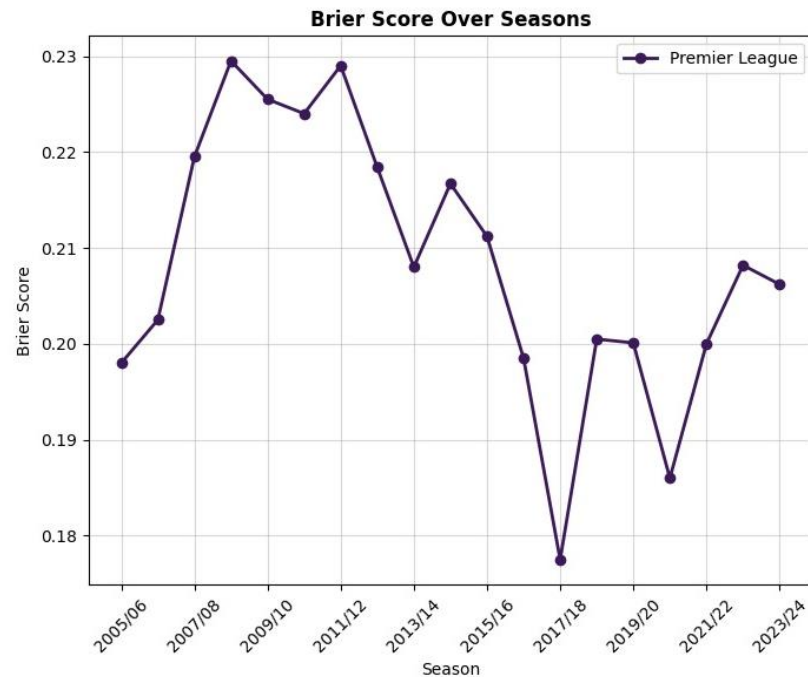
Bookmakers' Margin

Average Bookmaker Margin
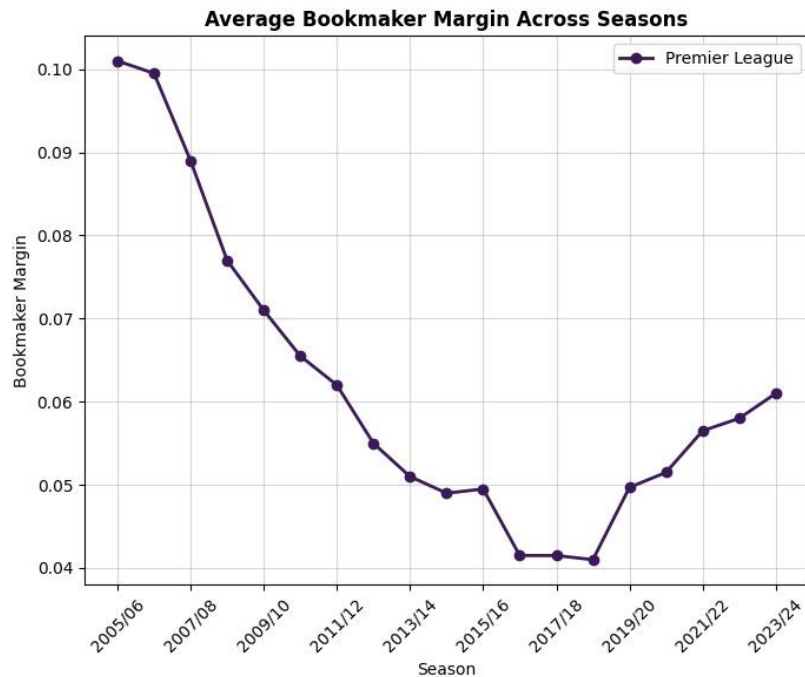
Accuracy of Probabilistic Predictions

Betting on Home Teams

Setup

- **Betting on Home Teams**: Exploit home bias

- **Strategy Explanation:**
  - Consists of betting on home teams for each match considered during the 19 seasons
  - Produces positive returns in 7 out of 19 seasons
  - Average return is **−2.44%**

Results

Season	Return
2005-06	2.82%
2006-07	-0.31%
2007-08	-10.22%
2008-09	-6.85%
2009-10	7.82%
2010-11	-0.20%
2011-12	-4.49%
2012-13	-11.61%
2013-14	-0.22%
2014-15	-2.99%
2015-16	-10.10%
2016-17	5.52%
2017-18	1.49%
2018-19	3.47%
2019-20	4.89%
2020-21	-15.86%
2021-22	-10.49%
2022-23	7.60%
2023-24	-6.64%
Average Return	**-2.44%**

Betting on Promoted Teams

Setup

- **Betting on Promoted Teams**: Focus on newly promoted teams

- **Strategies Results:**
  - Betting on promoted teams at home: -5.29%
  - Betting vs promoted teams at home: -5.70%
  - Betting vs promoted teams away: **2.46%** -> still below average bookmakers' margin

Results

Season	Return
2005-06	-9.07%
2006-07	4.23%
2007-08	27.13%
2008-09	1.93%
2009-10	9.20%
2010-11	-13.36%
2011-12	0.84%
2012-13	1.92%
2013-14	6.51%
2014-15	12.73%
2015-16	-1.59%
2016-17	14.51%
2017-18	-5.43%
2018-19	13.21%
2019-20	-4.20%
2020-21	-11.59%
2021-22	-1.58%
2022-23	-4.36%
2023-24	5.68%
Average Return	**2.46%**

Sentiment-Based Strategy

Setup

- **Sentiment-Based Strategy**: Betting on the most popular teams

- **Strategy Explanation:**
  - Popularity quantified by DiffAttend: strategy places bets only when the difference exceeds the 95th percentile
  - Astonishing returns in the first half of the sample, but only 1 out of the last 9 seasons was profitable
  - Average returns: **2.79%** -> still below average bookmakers' margin

Results

| Season	Return
2005-06	0.52%
2006-07	9.82%
2007-08	7.26%
2008-09	21.68%
2009-10	16.94%
2010-11	17.46%
2011-12	8.50%
2012-13	26.96%
2013-14	-5.88%
2014-15	-6.92%
2015-16	8.78%
2016-17	-29.70%
2017-18	-1.04%
2018-19	-4.94%
2021-22	-9.52%
2022-23	-1.46%
2023-24	-10.90%
Average Return	**2.80%**

Results and Findings

Results and Findings

Strategy Payoffs and Overpricing

- **Profitability Across Strategies:**
  - Betting on home teams: Negative long-term returns (-2.44%)
  - Betting against promoted teams (away): Slightly profitable (2.45%)
  - Sentiment-based strategy: Initially promising but declined post-2013

- **Biases Found**:
  - Underdogs are often overvalued away (favorite-longshot bias)
  - Popular teams had consistent overpricing pre-2013

Results

Predicted Home Win Probability
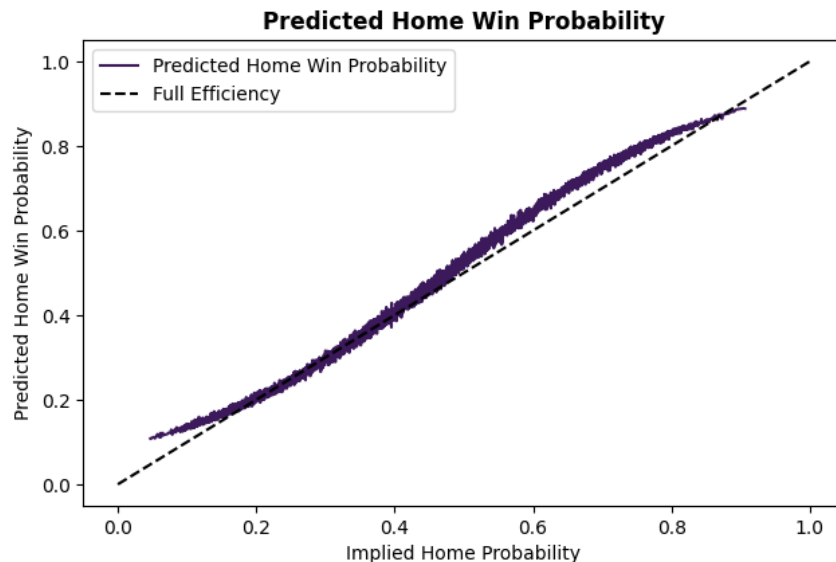
Challenges

Challenges

Patterns, Statistical Noise, & Limitations

- **Inconsistent Patterns:** Biases fluctuate seasonally

- **Impact of Statistical Noise:** Monte Carlo simulations revealed that apparent biases may result from random fluctuations

Monte Carlo Simulation

#Seasons	p=0.1	p=0.05	p=0.01
0	100.0	100.0	99.81
1	100.0	100.0	98.4
2	100.0	99.94	93.34
3	100.0	99.81	82.0
4	99.98	99.43	63.52
5	99.94	97.65	41.99
6	99.55	93.1	23.11
7	98.17	84.0	10.93
8	94.41	68.8	3.96
9	86.01	49.88	1.37
10	71.48	30.77	0.31
11	52.34	16.08	0.05
12	31.99	6.34	0.0
13	15.12	1.92	0.0
14	5.4	0.45	0.0
15	1.22	0.03	0.0
16	0.11	0.0	0.0

Conclusions

Conclusions

Key Takeaways

- Bookmakers' predictive power has increased through time on the back of compressed margins and thus a more competitive industry

- Betting markets are largely efficient over time

- There are temporary inefficiencies that bettors cannot reliably exploit

Future Research Ideas

- Expand to other leagues (e.g., Serie A, La Liga, etc.)

- Test advanced models like neural networks for predicting outcomes

- Expand research to other areas of betting such as political bets (outcome of elections, content of speeches)

Thank You for Your Attention!

Sources

Papers, Data & Other Materials

- Winkelmann D., Ötting M. and Deutscher C., "Betting Market Inefficiencies in European Football – Bookmakers' Mispricing or Pure Chance?", 2020
- Meier P.F., Flepp R. and Franck E., "Are sports betting market semistrong efficient? Evidence from the COVID-19 pandemic", Internation Journal of Sport Finance, 2021
- Angelini G. and De Angelis L., "Efficiency of online football betting markets", 2017
- Daunhawer I., Schoch D. and Kosub S., "Biases in the football betting market", 2017
- Direr A., "Are betting markets efficient? Evidence from European Football Championships", 2011
- Franke M., "Do market participants misprice lottery-type assets? Evidence from the European soccer betting market", 2020
- Statista Global Sports Betting Revenue Data