

Relatore:  
Carlo Ambrogio Favero

Tesi Magistrale di:  
Nicola Sita, matricola 3007204

Msc in International Management  
Class of 2020

*16<sup>th</sup> December 2020*



Università Commerciale  
Luigi Bocconi



# Evaluating the Home Advantage in NBA.

## Evidence from 2020



# OUTLINE

- 1. Context
- 2. Introduction on Sport Analytics
- 3. Literature Review on Home Advantage
- 4. Econometrics Theory
- 5. Regression Model
- 6. Results
- 7. Conclusions

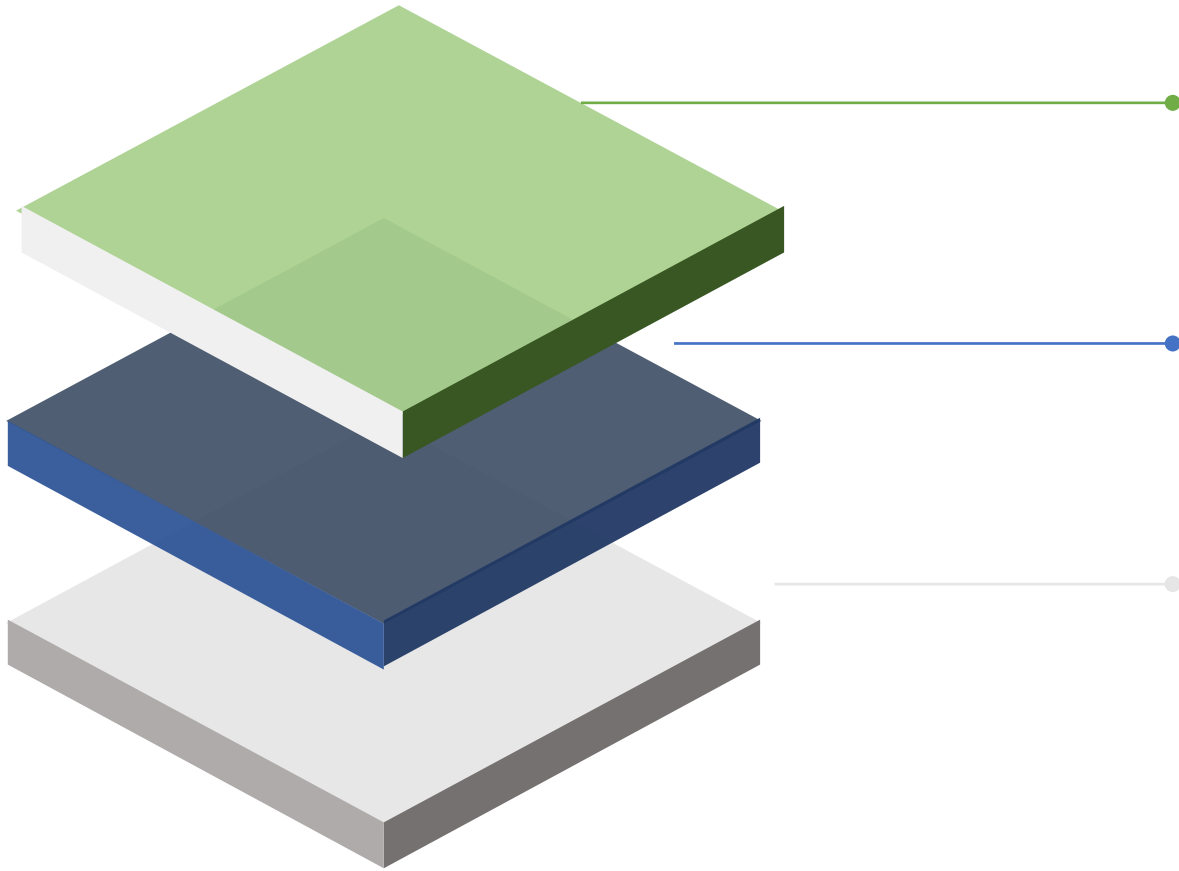


- **1. Context**

- 2. Introduction on Sport Analytics
- 3. Literature Review on Home Advantage
- 4. Econometrics Theory
- 5. Regression Model
- 6. Results
- 7. Conclusions

# Thesis Flow and Goal

---



## **Pandemic**

Due to pandemic, almost in all sport games, the presence of public was(/is) forbidden

## **NBA and COVID-19**

The NBA finished its season in a bubble (at Disneyworld Park in Orlando)

## **Exploit the Opportunity**

Unprecedented opportunity to compare these games with normal ones to assess the home advantage in NBA



- 1. Context
- **2. Introduction on Sport Analytics**
- 3. Literature Review on Home Advantage
- 4. Econometrics Theory
- 5. Regression Model
- 6. Results
- 7. Conclusions

# Sport Analytics in Basketball

The impact on basketball has been huge:

- **Every NBA franchise has its own data analysis department** in the front office
- The NBA itself organizes a **Hackaton** to discover data analyst talents
- Since 2009, in all the arenas, there is a **technology named "Player Tracking"**, used to record the players 25 times per second



## AVOIDING INJURIES

- Missing games when injury prone: «Load Management»
- Monitoring players constantly
- Players' diet



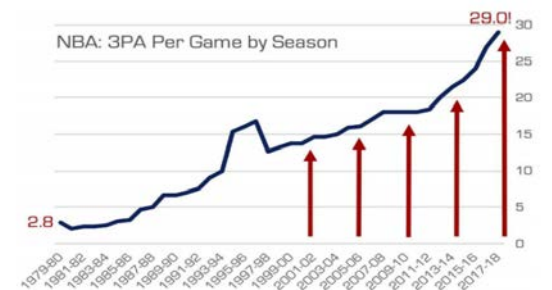
## SCOUTING

- Spread use of data analytics to pick players at Drafts
- Predict what kind of player a prospect can be
- Find the best fit



## WINNING STRATEGIES

- Most efficient solutions (3pt shot revolution/game strategies)





- 1. Context
- 2. Introduction on Sport Analytics
- **3. Literature Review on Home Advantage**
- 4. Econometrics Theory
- 5. Regression Model
- 6. Results
- 7. Conclusions

# Home Advantage: Definitions and Roots

Home Advantage is “the term used to describe the consistent finding that **home teams in sport competitions win over 50% of the games played** under a balanced home and away schedule”

**Table II.** The combined home winning percentages (weighted) for the major team sports

Sport	Number of studies	Number of games	HWP <sup>a</sup> (%)	HWP <sup>b</sup> (%)
Baseball	1	133 560		54.3
American football	5	2592	57.3	57.3
Ice hockey	5	5312	59.5	61.2
Basketball	9	13 686		64.4
Soccer	3	40 493	63.9	68.3

**HWP<sup>a</sup>** = home winning percentage (2 for a win, 1 for a draw and 0 for a loss); **HWP<sup>b</sup>** = home winning percentage (drawn games excluded).

## Game Location Factors

Characteristics that can explain the home advantage

### Main influencing factors:

- Crowd
- Learning
- Travel
- Rules

## Critical Psychological States

Influence based on psychological components

### “Actors” involved:

- Competitors
- Coaches
- Officials

## Critical Behavioural States

Influence based on behavioural components

### “Actors” involved:

- Competitors
- Coaches
- Officials

## Performance Outcomes

Outcomes of home advantage

### Three potential outcomes:

- Primary
- Secondary
- Tertiary



# Home Advantage: Methods

## Method 1, Pollard and Pollard (2005)

### 1. Home advantage as

Points obtained playing home/Total points achievable

### 2. Pro

Easy to compare different sports

### 3. Cons

Ignore the magnitude of performances

i.e.	Home	Away	Home Advantage
Team A	1-0 (W)	0-3 (W)	50%
Team B	3-0 (W)	0-1 (W)	50%

→ Same home advantage despite different outcomes

## Method 2, Marek and Vávra (2018)

### 1. Home advantage as

Difference of goals (points) scored, home-away

### 2. Pro

Assess real differences of performance

### 3. Cons

Goals scored can hide losing frequency

Team A	Game 1	Game 2	Game 3	Home advantage
Home	10-0 (W)	0-1 (L)	0-1 (L)	10-2 (goals) but 33% W
Away	0-1 (W)	0-1 (W)	2-0 (L)	2-2 (goals) but 67% W

→ At home team A seems better but it won less



- 1. Context
- 2. Introduction on Sport Analytics
- 3. Literature Review on Home Advantage
- **4. Econometrics Theory**
- 5. Regression Model
- 6. Results
- 7. Conclusions

# Theoretical Pillars

## EXPERIMENTS



Structure of econometrics experiments.

- 2020 games will be in the treatment group
- 2018/2019 games in the control one

## TWO METHODS



Same models with two methods:

- $Y_i = \text{home winning \%}$
- $Y_i = \text{differential}^1 \text{ home} - \text{differential away}$

## VARIABLES



Variables used to check home advantage influences:

- The bubble (absence of public)
  - The round of Playoff

## SET OF MODELS



$$Y_i = \beta_0 + \beta_1 2018 + \beta_2 2019 + \beta_3 2020 + u_i$$

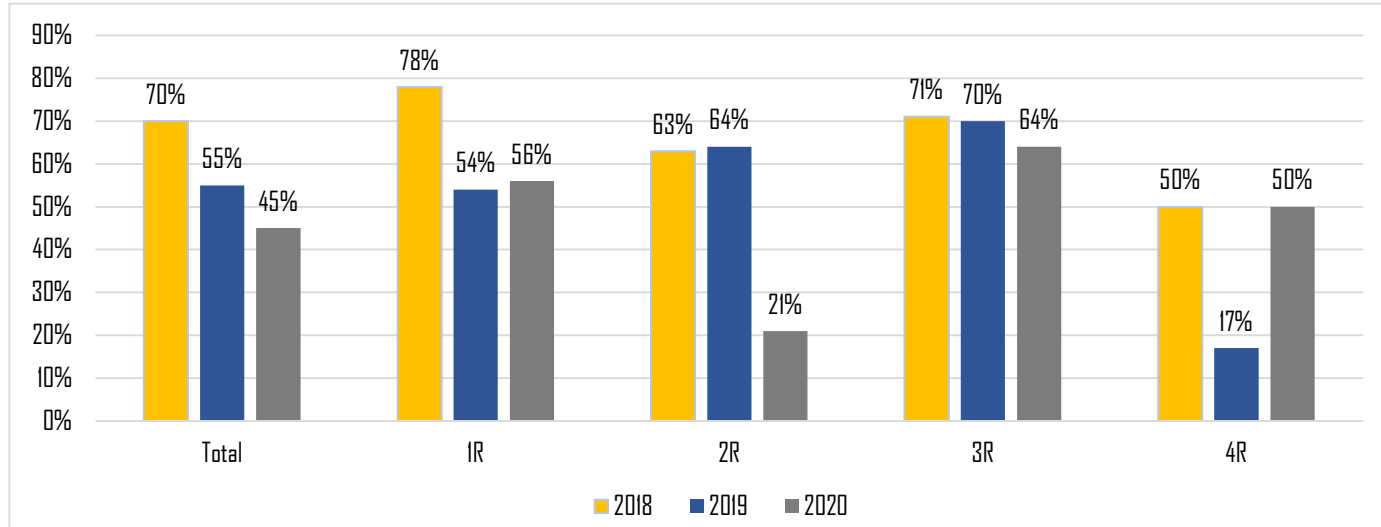
$$Y_i = \beta_1 \text{Round1} + \beta_2 \text{Round2} + \beta_3 \text{Round3} + \beta_4 \text{Round4} + \beta_5 \text{Round1}_{2020} + \beta_6 \text{Round2}_{2020} + \beta_7 \text{Round3}_{2020} + \beta_8 \text{Round4}_{2020} + u_i$$

<sup>1</sup>: differential= points scored – points allowed



- 1. Context
- 2. Introduction on Sport Analytics
- 3. Literature Review on Home Advantage
- 4. Econometrics Theory
- **5. Regression Model**
- 6. Results
- 7. Conclusions

# Descriptive Analysis



## WINNING % PER ROUND

- In 2020 for the first time the winning % of the home team has been < 50%
- In the Second Round, this % for 2020 was 21% compared to 63% and 64% of previous years

Playoff 2020	W%	FG %	3P %	FT %	TRB	AST	PTS	ORtg
Home	45%	46%	36%	80%	43.35	22.82	110.06	112.59
Away	55%	46%	36%	78%	42.95	22.92	109.06	111.61

## GAME STATISTICS HOME VS AWAY

- As we can see there is almost no difference in statistics between the home and the away team
- Even if the home team won 45% of games, it scored one more point

# Models Flow

There will be two sets of models, each one with the same 5 regressions and with the only difference in the dependent variable  $Y_i$



The five regression can be grouped into three subsets: basic scenario, years comparison and rounds comparison

## Regression groups

### Basic scenario

**Goal:**

compute the home advantage without variables

**Structure:**

```
lm1 <- lm(NBAdata$DIFF ~ 1)
```

### Years comparison

**Goal:**

assess the home advantage for every year (and 2020 vs past)

**Structure:**

```
lm2_1 <- lm(NBAdata$DIFF ~ 1 + NBAdata$dum2020)
```

### Rounds comparison

**Goal:**

assess the home advantage for every round (and 2020 specific)

**Structure:**

```
lm3 <- lm(NBAdata$DIFF ~ NBAdata$dumROUND1 +  
NBAdata$dumROUND2 + NBAdata$dumROUND3 +  
NBAdata$dumROUND4 - 1)
```



- 1. Context
- 2. Introduction on Sport Analytics
- 3. Literature Review on Home Advantage
- 4. Econometrics Theory
- 5. Regression Model
- **6. Results**
- 7. Conclusions

# Results Set of Models 1; Y = Margin

Model	Goal	Main Results (with 95% confidence interval)	Takeaway
Lm1 (Y = margin)	Define general home advantage	<b>Intercept: 3.6842 ± 0.1158</b>	<b>Between 2018 and 2020, home teams won on average by 3.68 pts</b>
Lm2 (Y = margin)	Define home advantage for each year	2018: 6.280 ± 0.196 2019: 3.805 ± 0.196 <b>2020: 1.000 ± 0.195</b>	<b>In 2020, the margin was much lower than in 2018/2019</b>
Lm2_1 (Y = margin)	Analyse the impact of 2020 against previous years	Intercept: 5.043 ± 0.139 <b>2020: -4.043 ± 0.239</b>	<b>In 2020, the home teams lost around 4 points in margin</b>
Lm3 (Y = margin)	Define the home advantage for every round	Round 1: 4.3203 ± 0.157 Round 2: 1.5588 ± 0.216 Round 3: 7.5429 ± 0.301 <b>Round 4: -0.8125 ± 0.445</b>	<b>The Finals were the only round with a negative home advantage</b>
Lm4 (Y = margin)	Compute the impact of 2020 on Rounds	Round 1_2020: <b>-2.355 ± 0.332</b> Round 2_2020: <b>-7.818 ± 0.448</b> Round 3_2020: <b>-6.360 ± 0.643</b> Round 4_2020: 5.833 ± 0.912	<b>Playing in the bubble had a negative effect on all round (except Finals)</b>



# Results Set of Models 2; Y = Winning %

Model	Goal	Main Results (with 95% confidence interval)	Takeaway
Lm5 (Y = winning %)	Define general home advantage	<b>Intercept: 58.30% ± 0.4%</b>	<b>The home advantage found is 58.30%</b>
Lm6 (Y = winning %)	Define home advantage for each year	2018: 70.73% ± 0.67% 2019: 56.10% ± 0.67% <b>2020: 48.19% ± 0.67%</b>	<b>At 48.19%, the 2020 is the only year with a negative home advantage</b>
Lm6_1 (Y = winning %)	Analyse the impact of 2020 against previous years	Intercept: 63.41% ± 0.48% <b>2020: -15.22% ± 0.82%</b>	<b>Compared to previous years, in 2020 home teams won 15% less games</b>
Lm7 (Y = winning %)	Define the home advantage for every round	Round 1: 63.28% ± 0.54% <b>Round 2: 48.53% ± 0.74%</b> <b>Round 3: 68.57% ± 1.03%</b> Round 4: 37.50% ± 1.52%	<b>Round 2 and 4 have values lower than 50%. Instead, Round 3 is close to 70%</b>
Lm8 (Y = winning %)	Compute the impact of 2020 on Rounds	Round 1_2020: -5.59% ± 1.13% <b>Round 2_2020: -42.80% ± 1.52%</b> Round 3_2020: -7.20% ± 2.18% Round 4_2020: 20.00% ± 3.09%	<b>The main difference is given by Round 2_2020 where the home advantage was lower by almost 43%</b>



- 1. Context
- 2. Introduction on Sport Analytics
- 3. Literature Review on Home Advantage
- 4. Econometrics Theory
- 5. Regression Model
- 6. Results
- **7. Conclusions**

# Main Findings

## Huge Impact of Public

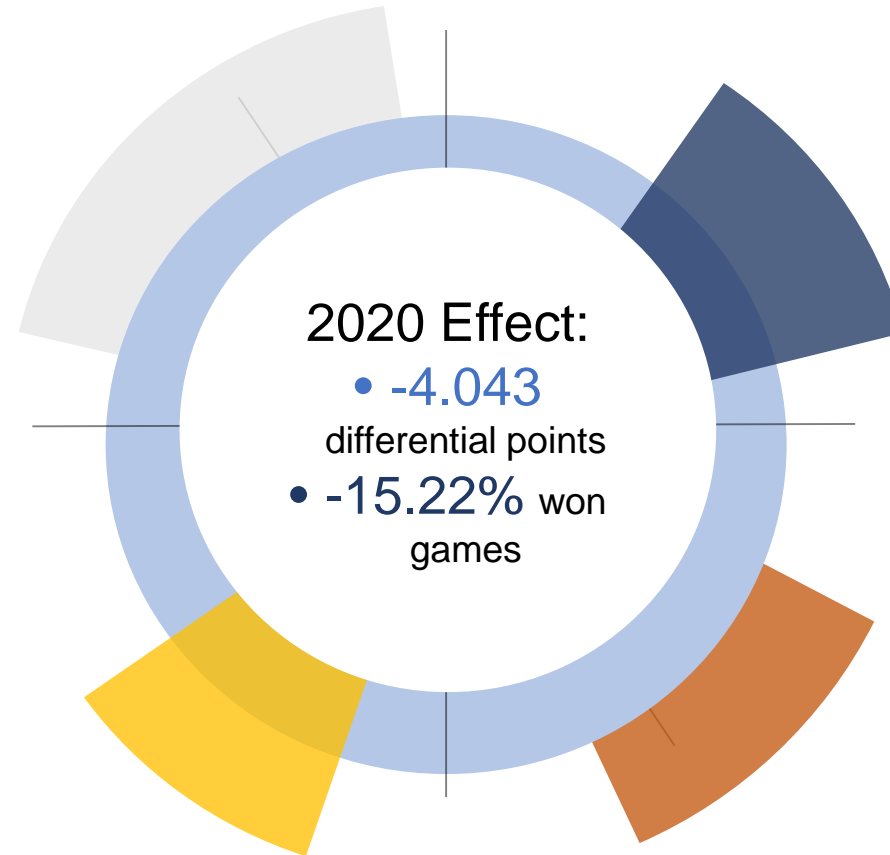
Compared with games with public, in the bubble home teams:

- Had a lower differential by **4.043** points
- Won **15.22%** fewer games

## In line with literature

Basketball home advantage in literature was between **60-64%**

Just considering 2018/2019 games, the estimated home advantage has been **63.42%**



## Home Advantage Factors

Considering the initial framework, the absent factors in the bubble were:

- **Crowd**
- **Familiarity with the venue**
- **Travel**

## Round Impact

- Round 3 was the one with the highest home advantage (68.57%)
  - **In the Finals home «disadvantage» (37.50%)**

# CONCLUSION



**Situation:** due to COVID-19, in NBA, for the first time games with completely null home advantage



**Opportunity:** computing home advantage through comparison between normal games and 2020 ones



**Results** satisfied hypothesis: home advantage exists! It increase the likelihood of winning by 15%



**Thank you!**

Questions?