

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Informatica

**Oltre i Big Man:
l'evoluzione del gioco NBA attraverso
l'analisi statistica del tiro da tre punti**

Relatore:
Chiar.mo Prof.
Marco Roccetti

Presentata da:
Francesco Saverio
Beccafichi

I Sessione
Anno Accademico 2023/2024

Sommario

Negli ultimi decenni, il gioco NBA ha subito una trasformazione radicale. I giocatori NBA più popolari sono passati da essere i dominatori del pitturato come Shaquille O'Neal, David Robinson, Tim Duncan e Kareem Abdul-Jabbar, che erano fondamentali per le tattiche di gioco, a giocatori più versatili e dinamici come Stephen Curry, James Harden e Damian Lillard, che prediligono il gioco perimetrale e il tiro dalla lunga distanza. Infatti, in questi ultimi periodi, è sempre più comune vedere tiri da centrocampo durante le partite, non solo allo scadere del tempo regolamentare.

L'obiettivo di questa tesi è analizzare le tendenze NBA che si sono sviluppate nel corso delle varie stagioni, ponendo il focus sull'impatto del tiro da tre punti nel principale campionato di basket mondiale. Lo studio mira a evidenziare come l'aumento del tiro da tre punti abbia influenzato vari aspetti del gioco, in modo positivo o negativo. Questa ricerca esamina i dati delle stagioni NBA, analizzando nello specifico il numero di tiri da due punti e da tre punti effettuati ogni anno. Questi dati verranno utilizzati per determinare se influenzano aspetti del gioco, come il numero di punti segnati per partita e la percentuale di vittorie che una squadra può raggiungere in una stagione.

L'analisi statistica verrà condotta utilizzando tecniche di data science per identificare pattern e tendenze significative. I risultati di questa analisi hanno lo scopo di fornire informazioni utili sia per gli appassionati di basket sia per i professionisti del settore, ma soprattutto di confermare statisticamente la tendenza delle squadre NBA a tirare sempre più da tre punti.

La NBA è da sempre considerata la massima espressione della pallacanestro a livello mondiale, poiché raggruppa i migliori talenti tra giocatori, allenatori e dirigenti. Questo la rende un modello di riferimento per tutti i campionati di pallacanestro nel mondo, che cercano costantemente di emularne le strategie, sia a livello di marketing che a livello di tattiche.

Negli ultimi dieci anni, i modelli di gioco delle squadre NBA si sono affidati sempre di più all'utilizzo del tiro da tre punti, rivoluzionando gli schemi che permettevano di avere stagioni di successo nei primi anni 2000. L'obiettivo di questo studio è analizzare se questa tendenza influenza in modo statisticamente significativo alcuni aspetti del gioco, mettendo in luce l'evoluzione delle tattiche che sfruttano sempre di più il tiro dalla lunga distanza nel massimo campionato di basket mondiale.

Questo studio cerca di analizzare come l'aumento del tiro da tre punti influenzi, in modo positivo o negativo, aspetti cruciali del gioco. In particolare, i dati verranno utilizzati per determinare l'influenza su due aspetti specifici: il numero di tiri da tre punti tentati e i punti segnati per partita di una squadra; e il numero di tiri da tre punti tentati e il numero di vittorie che una squadra può raggiungere in una stagione.

I risultati di questa ricerca puntano a confermare o smentire, a livello statistico, ciò che è evidente osservando le partite di basket dagli anni 2000 ad oggi. Questa analisi fornirà una comprensione approfondita dell'impatto del tiro da tre punti sulle dinamiche di gioco e sul successo delle squadre NBA.

Contents

1	Introduzione	1
1.1	Revisione della letteratura	2
1.2	Motivi della ricerca	4
2	Dati e Metodi	7
2.1	Dati	8
2.2	Metodologie	15
2.2.1	Serie temporali	17
2.2.2	Rilevamento dei Changepoint	19
2.2.3	Test di Mann-Kendall	22
2.2.4	Test della stazionarietà	23
2.2.5	Analisi della Correlazione	26
2.2.6	Test di causalità di Granger	27
2.2.7	Regressione Lineare	29
3	Risultati	31
3.1	Rilevamento dei changepoint per i tiri effettuati in NBA	32
3.2	Rilevamento di Trend nei tiri tentati	38
3.3	Correlazione tra tiri tentati e media punti a partita	42
3.3.1	Pearson Correlation	43
3.3.2	Spearman Correlation	44
3.3.3	Granger causality	47
3.4	Predizione vittorie	48

4 Conclusioni	51
4.1 Limiti dello studio	53
4.2 Conclusioni	54
4.3 Lavori Futuri	54
Bibliography	56

List of Figures

2.1	Dateset finale	11
2.2	Distribuzione dei tiri tentati.	13
2.3	Distribuzione dei tiri da due e tre punti	14
2.4	Trend tiri da tre punti tentati e realizzati con relativa media. .	15
2.5	Trend tiri da due punti tentati e realizzati con relativa media.	16
2.6	Serie temporali di tiri da due e tre punti per stagione.	18
3.1	Rilevamento dei changepoint nel numero di tiri da tre punti effettuati dal 1996	33
3.2	Deviazione standard e media dei tiri da 3 punti tentati.	34
3.3	Rilevamento dei changepoint nel numero di tiri da tre punti effettuati dal 1996.	36
3.4	Deviazione standard e media dei tiri da 2 punti tentati.	37

List of Tables

3.1	Probabilità di un changepoint tra il 1996 e 2024	33
3.2	Probabilità di un changepoint tra il 1996 e 2024	36
3.3	Risultati del test Mann-Kendall sui tiri da tre punti tentati.	38
3.4	Risultati del test ADF per i tiri da tre punti tentati.	39
3.5	Risultati del test KPSS per i tiri da tre punti tentati.	39
3.6	Risultati del test Mann-Kendall sui tiri da due punti tentati.	40
3.7	Risultati del test ADF per i tiri da due punti tentati.	41
3.8	Risultati del test KPSS per i tiri da due punti tentati.	41
3.9	Risultato della Cross-Correlation tra i tiri da tre punti tentati e i punti segnati per partita.	43
3.10	Risultato della Cross-Correlation tra i tiri da due punti tentati e i punti segnati per partita.	43
3.11	Risultato della correlazione di Spearman tra i tiri da tre punti tentati e i punti segnati per partita prima del 2009/10.	45
3.12	Risultato della correlazione di Spearman tra i tiri da tre punti tentati e i punti segnati per partita dopo 2009/10.	45
3.13	Risultato della correlazione di Spearman tra i tiri da due punti tentati e i punti segnati per partita prima del 2009/10.	46
3.14	Risultato della correlazione di Spearman tra i tiri da due punti tentati e i punti segnati per partita dopo il 2009/10.	46
3.15	Risultati del test di Granger Causality tra i tiri da tre punti tentati e i punti segnati per partita.	47

3.16 Risultati del test di Granger Causality tra i tiri da due punti tentati e i punti segnati per partita.	48
3.17 Risultati MSE e R2 Score per la regressione lineare.	48
3.18 Coefficienti della regressione lineare	49

Chapter 1

Introduzione

La NBA è da sempre considerata il campionato di basket più competitivo al mondo; per questo motivo, chi vince il titolo viene definito "World Champion" nonostante si tratti di una competizione di livello nazionale. Come in tutti gli sport, anche la NBA ha subito diverse influenze tattiche che hanno portato le squadre ad adattarsi agli stili di gioco più efficienti. È evidente come il gioco NBA sia cambiato radicalmente dagli anni 2000 fino ad oggi. Nei primi anni del nuovo millennio, i giocatori più importanti all'interno delle squadre, ovvero coloro a cui venivano affidate le maggiori responsabilità nei sistemi offensivi, erano i centri: atleti di statura ben oltre la media e fisici possenti, che dominavano le aree intorno al canestro. Questo tipo di gioco portava a un ritmo di partita molto basso, dove tutti i giocatori in campo venivano coinvolti, ma l'obiettivo finale di uno schema era mettere la palla in post basso e dare priorità alle doti offensive del proprio centro per tentare di fare canestro. Esempi emblematici sono i Los Angeles Lakers di Shaquille O'Neal, che riuscirono nell'impresa di vincere tre titoli NBA consecutivi, i San Antonio Spurs di Tim Duncan, riconosciuti come una delle squadre più vincenti della storia NBA, e i Boston Celtics di Kevin Garnett. Al contrario, nella pallacanestro moderna, viene prediletto il gioco veloce, dove l'obiettivo principale di un attacco è aumentare il ritmo di gioco. Per rendere questo possibile, alcune squadre basano le loro tattiche su situazioni

di uno contro uno. È sempre più frequente vedere durante le azioni NBA situazioni di *pick and pop*, dove il playmaker sfrutta un blocco del pivot e attacca il vantaggio creato da questa situazione. Gli altri componenti della squadra si dispongono sulla linea dei tre punti pronti a ricevere uno scarico per tentare un tiro. Questo schema mette la difesa di fronte a una scelta: subire un canestro facile in appoggio vicino al ferro oppure far tirare un avversario da tre punti. Questo ha portato a rendere il gioco estremamente rapido, esasperando concetti come il "*Seven Seconds or Less*"[1] proposto da Mike D'Antoni con i Phoenix Suns, in cui le azioni offensive dovevano durare sette secondi o meno. Le azioni iniziavano con un *pick and roll* e il primo vantaggio generato doveva essere sfruttato per tentare di fare canestro.

Un fattore determinante in questa evoluzione è stato l'influenza dei cosiddetti *Splash Brothers*[2], che a partire dal 2012 hanno rivoluzionato il modo di interpretare la pallacanestro, ponendo il tiro da tre punti al centro del proprio sistema offensivo.

Uno degli esponenti principali di questa rivoluzione è Stephen Curry, che ha sfruttato questo approccio per infrangere numerosi record relativi ai tiri da tre punti. Curry ha superato il record di Ray Allen per il maggior numero di tiri da tre punti realizzati in carriera nel 2021, consolidando la sua posizione come uno dei migliori tiratori di sempre.

Il duo composto da Curry e Klay Thompson, il secondo membro degli *Splash Brothers*, integrato nel sistema di gioco adottato da Steve Kerr nei Golden State Warriors, ha permesso a questa franchigia di dominare il campionato dal 2014 al 2022. Durante questo periodo, i Warriors hanno vinto quattro titoli NBA e hanno stabilito il miglior record di sempre in una stagione NBA (73-9).

1.1 Revisione della letteratura

La letteratura esistente sull'evoluzione del gioco NBA mostra una chiara transizione dal predominio dei centri all'enfasi sul tiro da tre punti. Nell'ar-

articolo scritto da Lawyer[3], viene riportato come i centri NBA abbiano avuto molta importanza non solo nei primi anni 2000, ma anche nell’NBA degli anni ’80, dove le squadre venivano costruite intorno ai pivot, che avevano il compito di difendere il canestro e garantire tiri ad alta percentuale in attacco. Viene evidenziato come, da quando Shaquille O’Neal venne premiato Most Valuable Player (MVP) nel 2000, nessun altro centro sia stato in grado di ricevere questo premio fino a Nikola Jokic nel 2021, indicando un periodo di 21 anni in cui nessun centro è stato premiato a livello individuale per aver portato una squadra alla vittoria del titolo NBA. L’articolo sottolinea come i centri si siano dovuti adattare ai cambiamenti del gioco, con Jokic che ha aggiunto al suo bagaglio tecnico molte più abilità perimetrali rispetto a O’Neal, rendendo comune vedere il centro serbo palleggiare per tutto il campo o tirare da tre punti dopo un blocco di un compagno.

Un altro articolo che evidenzia il cambiamento nel ruolo del centro è quello pubblicato sul *Journal of Sport Analytics*[4], che si concentra su come l’introduzione del tiro da tre punti abbia inizialmente diminuito il valore offensivo dei centri e come questi ultimi abbiano dovuto adattare il loro stile di gioco per rimanere efficaci nell’era dei tre punti. L’articolo analizza come il valore dei centri sia diminuito dopo l’introduzione del tiro da tre punti, per poi tornare ai livelli precedenti solo negli ultimi anni.

Il paper di Sanberg[5] si focalizza sulla rivoluzione del tiro da tre punti in NBA, discutendo i trend e le correlazioni nel tiro da tre punti e sottolineando come i tentativi da tre punti siano aumentati nel corso degli anni, influenzando la strategia di gioco complessiva e il successo delle squadre. Viene evidenziato come, nonostante l’aumento dei tiri da tre punti, le percentuali siano rimaste stabili, con esempi specifici come il successo di Stephen Curry e dei Houston Rockets, che hanno fatto ampio affidamento su questo aspetto del gioco.

Il successo di Stephen Curry e dei Golden State Warriors viene riconosciuto da Williams e White[6] come una delle principali cause dell’ampio utilizzo del tiro da tre punti in NBA. Analizzando l’abilità di Curry nel tiro da tre

punti, viene mostrato come questo influisca sul sistema di gioco della squadra e sui risultati ottenuti. Alcuni analisti NBA e specialisti hanno sviluppato modelli per ottimizzare le strategie offensive dei giocatori, come descritto da Ehrlich[7], che ha creato un nuovo tipo di tabella di tiro utilizzando un modello additivo generalizzato per stimare l'abilità di tiro di un giocatore, considerando anche i tiri sbagliati che comportano un fallo di tiro e i punteggi dei tiri liberi successivi.

1.2 Motivi della ricerca

Lo studio presentato si propone di dimostrare a livello statistico se l'uso massiccio del tiro da tre punti porti maggiori vantaggi alle squadre NBA rispetto ai tiri da due punti. Questo studio è motivato dalla necessità di comprendere meglio le dinamiche che hanno portato al crescente predominio del tiro da tre punti nel basket professionistico e di analizzare i vantaggi strategici che questa tattica offre alle squadre.

Si intende comprendere la distribuzione del tiro da tre punti nelle varie stagioni NBA per verificare un eventuale aumento significativo e il conseguente cambiamento nello stile di gioco delle squadre. Sarà analizzata l'efficienza statistica di questo cambiamento, dimostrando se è realmente vantaggioso sfruttare così tanto questo aspetto del gioco.

Il tiro da tre punti può rivelare informazioni interessanti su come le squadre adattano la costruzione dei roster in base alle tattiche di gioco. Pertanto, sarà effettuato un confronto su come i centri delle squadre abbiano modificato i loro stili di gioco e le loro abilità per rimanere rilevanti nell'NBA moderna. La tesi si concentrerà sulle seguenti domande di ricerca:

- Analizzare la correlazione tra il numero di tiri da tre punti tentati e il numero di punti segnati per partita.
- Analizzare con la statistica inferenziale le tattiche che permettono di aumentare la percentuale di vittorie di una squadra.

- Analizzare se nella NBA moderna è più efficiente tirare da due punti o tirare da tre punti.

Chapter 2

Dati e Metodi

Le metodologie e i dati esaminati in questo studio prevedono un'analisi completa del numero di tiri da tre punti tentati da tutti i giocatori NBA, dalla stagione 1996/97 alla stagione 2023/2024. Lo studio è diviso in tre parti principali:

- **Analisi Quantitativa e Qualitativa:** Viene effettuata un'analisi statistica del numero di tiri da due punti e da tre punti tentati dai giocatori NBA, ricercando eventuali tendenze e modelli nel corso delle stagioni selezionate.
- **Analisi di Correlazione e Causalità:** I dati vengono utilizzati per determinare l'influenza delle due tipologie di tiro sul numero di punti segnati per partita. Lo stesso tipo di analisi viene effettuata in due serie temporali differenti, in modo da confrontare se effettivamente c'è stato un cambiamento e se questo cambiamento ha portato a un miglioramento.
- **Previsione delle Vittorie NBA:** I dati raccolti saranno impiegati per determinare quali variabili del gioco hanno un impatto significativo sulla percentuale di vittorie di una squadra NBA al termine della stagione.

2.1 Dati

I dati utilizzati per questa tesi sono stati ottenuti da diverse fonti. Le statistiche NBA riguardanti le stagioni giocate dal 1996 al 2024 sono state reperite da *basketball-reference*, un sito web di proprietà di Sport Reference, con sede principale a Philadelphia. Il sito contiene informazioni dettagliate riguardanti il mondo della NBA, suddivise per stagioni, con un'ulteriore divisione tra *regular season* e *playoff*. All'interno del sito è possibile visualizzare gli storici di tutte le stagioni con le relative statistiche, che vengono ulteriormente suddivise in differenti tabelle in base al tipo di visualizzazione scelto (per partita, totali, su cento possesi).

Le stagioni NBA analizzate riguardano il periodo dal 1996 al 2024 ed è stata considerata solamente la *regular season*, in modo da includere tutte le trenta squadre facenti parte della lega e non solo le migliori sedici che partecipano alla post-season. Un altro motivo per cui è stata scelta la regular season è il fatto che alcune squadre NBA tendono ad adottare una strategia mirata alla sconfitta, poiché all'inizio di ogni stagione la lega organizza un draft in cui le squadre possono selezionare i migliori prospetti al mondo per aggiungerli al proprio roster. Una squadra con un record negativo ha più probabilità di ottenere scelte migliori alla lottery[8].

Per quanto riguarda i dati utilizzati, sono state selezionate le tabelle contenenti le statistiche complessive di ogni squadra partecipante nelle stagioni considerate. Attraverso un processo di web scraping, è stata implementata la funzione *get_tables* per ottenere tutte le tabelle desiderate. La funzione utilizza la libreria *requests*[9] per scaricare il contenuto della pagina specificata nella variabile *url*. Successivamente, mediante l'uso di *BeautifulSoup*[10], è possibile analizzare e manipolare il codice HTML. In particolare, viene individuato l'elemento *div* con l'ID '*div_totals-team*' per estrarre la tabella desiderata. Prima di essere salvata in un file *.csv*, viene aggiunta una colonna '*SEASON*' alla tabella, indicando l'anno specificato dalla variabile *year*. In una fase successiva, verrà ulteriormente elaborata includendo anche l'anno successivo, così da ottenere la stagione di riferimento. Nel caso di errori

durante questo procedimento, viene generata una notifica tramite stampa.

```
def get_tables(url, year):
    response = requests.get(url)

    if response.status_code == 200:
        soup = BeautifulSoup(response.content, 'html.parser')

        total_stats_tables = soup.find_all('div', id='div_totals-team')

        for idx, table in enumerate(total_stats_tables, start=1):
            df = pd.read_html(str(table))[0]
            df['SEASON'] = str(year)
            file_name = f'total_stats_{year}.csv'
            df.to_csv(file_name, index=False)

            print(f'Tabella "Total Stats" per la stagione {year}, tabella
                  ↪ {idx} salvata su {file_name}')
        else:
            print(f"Errore {response.status_code} durante il recupero della
                  ↪ pagina {url}")
```

La colonna *SEASON* è stata inclusa per consentire l'identificazione della stagione in cui una particolare squadra ha generato le statistiche. Questa aggiunta è fondamentale poiché, utilizzando la funzione *load_csv*, tutti i file CSV precedentemente scaricati verranno consolidati in un unico dataset. Questo approccio permette di avere tutte le statistiche necessarie in un unico archivio dati, semplificando ulteriormente l'analisi e la manipolazione dei dati.

```
def load_csv(dir):
    file = os.listdir(dir)
    csv_files = [f for f in file if f.endswith('.csv')]
    dfs = []
    for f in csv_files:
        df = pd.read_csv(os.path.join(dir, f))
        dfs.append(df)
    df = pd.concat(dfs, ignore_index=True)
    return df
```

Il dataset utilizzato è stato arricchito con statistiche avanzate che riguardano aspetti del gioco come il ritmo di gioco e l'efficienza offensiva e difensiva di una squadra. Queste statistiche avanzate sono state ottenute dal sito Basketball-Reference (*basketball-reference*), dove sono disponibili tutte le statistiche avanzate raccolte durante le stagioni NBA. Utilizzando il medesimo metodo di web scraping descritto in precedenza, le statistiche sono state raccolte e salvate in file *.csv* separati per ogni stagione, per poi essere unificate in un unico dataset. Le statistiche avanzate utilizzate in questa ricerca sono le seguenti:

- **Pace**: indica la quantità di possessi che una squadra gioca per ogni partita, fornisce informazioni sul ritmo del gioco di una squadra NBA. Viene calcolata con la formula:

$$\text{Pace} = 48 \times \frac{\text{Team Possessions} + \text{Opponent Possessions}}{2 \times \left(\frac{\text{Team Minutes Played}}{5}\right)}$$

- **Offensive Rating (ORtg)**: indica il numero di punti che una squadra sengan in 100 possessi. Fornisce informazioni sull'efficienza dell'attacco di una squadra. Viene calcolata con la formula:

$$\text{ORtg} = \frac{\text{Points Scored}}{\text{Total Possessions}} \times 100$$

- **Defensive Rating (DRtg)**: indica il numero di punti che una squadra concede agli avversari per 100 possessi. Fornisce informazioni sull'efficienza difensiva di una squadra. Viene calcolata con la formula:

$$\text{DRtg} = \frac{\text{Opponent Points Scored}}{\text{Opponent Possessions}} \times 100$$

Per completare la raccolta dei dati richiesti, è stato necessario ottenere i record di vittorie e sconfitte di ogni squadra durante la stagione. Queste informazioni sono state reperite da altre tabelle sul sito di Basketball-Reference (*basketball-reference*), suddivise tra la Eastern e Western Conference. Per integrare queste tabelle nei dati già raccolti, è stato necessario unirle verticalmente, garantendo così l'inclusione dei record di ogni squadra per ogni

stagione. Come parte di questa operazione, è stata aggiunta la colonna *SEASON* per indicare l'anno di riferimento, necessario per facilitare il processo di fusione dei due dataset.

Successivamente, il dataset è stato ripulito da tutte le righe superflue contenute nelle tabelle del sito, e sono stati controllati eventuali valori nulli. Inoltre, sono stati rimossi tutti gli asterischi, utilizzati per indicare le squadre partecipanti ai playoff. Questo è stato fatto al fine di poter raggruppare le statistiche per squadra, evitando ambiguità dovute alla presenza degli asterischi.

Infine, come anticipato, i dataset sono stati uniti tramite le colonne *Team* e *SEASON*, al fine di fornire tutte le statistiche necessarie per lo studio. Una volta completato il processo di raccolta dei dati, il dataset finale si presentava nella forma mostrata in Figura 2.1.

	SEASON	Team	W	L	Pace	ORtg	DRtg	GP	3P	3PA	2P	2PA	PTS	PPG
0	1997/98	Chicago Bulls	69.0	13.0	90.0	114.4	102.4	82	523	1403	2754	5520	8458	103.15
1	1997/98	Utah Jazz	64.0	18.0	90.0	113.6	104.0	82	334	902	2797	5315	8454	103.10
2	1997/98	Seattle SuperSonics	57.0	25.0	90.3	111.2	102.7	82	559	1583	2436	4832	8274	100.90
3	1997/98	Atlanta Hawks	56.0	26.0	86.8	108.5	102.3	82	659	1834	2153	4473	7774	94.80
4	1997/98	Miami Heat	61.0	21.0	88.2	106.8	100.6	82	678	1865	2144	4370	7776	94.83
...
792	1999/00	New Jersey Nets	16.0	34.0	91.5	99.5	103.6	50	225	679	1466	3481	4569	91.38
793	1999/00	Denver Nuggets	14.0	36.0	90.5	103.1	110.4	50	302	922	1379	3067	4674	93.48
794	1999/00	Vancouver Grizzlies	8.0	42.0	89.0	98.4	107.9	50	148	453	1495	3385	4443	88.86
795	1999/00	Los Angeles Clippers	9.0	41.0	89.5	100.0	109.7	50	214	668	1497	3339	4519	90.38
796	1999/00	Chicago Bulls	13.0	37.0	88.1	92.4	103.0	50	177	612	1362	3225	4095	81.90

Figure 2.1: Dataset finale

Le statistiche registrate delle squadre NBA sono state poi normalizzate in base al numero di partite svolte in ogni stagione. Questa scelta è stata dettata da diversi motivi.

Innanzitutto, dalla stagione 2004/05 l'NBA ha deciso di espandersi e passare da ventinove a trenta squadre con l'introduzione dei Charlotte Bobcats. Questo cambiamento ha comportato un maggior numero di partite giocate per stagione, rendendo necessaria una normalizzazione per mantenere la coerenza dei dati e facilitare un'analisi comparativa accurata.

In secondo luogo, durante il periodo di tempo preso in esame, ci sono stati due scioperi dei giocatori NBA che hanno portato alla riduzione del numero di partite giocate in due diverse stagioni. La stagione 1998/99 è stata ridotta a 50 partite e la stagione 2011/12 a 66 partite a causa di questi scioperi, influenzando significativamente il numero totale di partite giocate in quelle stagioni.

Infine, la pandemia di COVID-19 ha avuto un impatto notevole sul calendario NBA. La stagione 2019/20 è stata interrotta il 12 marzo 2020 e ripresa il 30 luglio 2020 con un formato ridotto che ha coinvolto solo le squadre con la possibilità di accedere ai playoff. Questo slittamento ha causato anche un ritardo nella ripresa della stagione successiva, portando le squadre NBA a giocare solo 72 partite nella stagione 2020/21.

Per mostrare la distribuzione dei tiri effettuati dalle squadre NBA, i dati sono stati suddivisi in due gruppi in base alle tipologie di tiro: tiri da due punti e tiri da tre punti. Per ogni stagione, è stato riportato il numero totale di tiri da due punti e da tre punti tentati dalle squadre NBA. Come mostrato nella Figura 2.2, l'utilizzo del tiro da tre punti è quasi raddoppiato nel corso delle stagioni, mentre è evidente una diminuzione nell'utilizzo del tiro da due punti.

Per comprendere se questo fenomeno è stato causato da una singola squadra o se rappresenta una tendenza generale adottata in NBA, sono state selezionate cinque stagioni differenti, distanziate equamente nel tempo, per osservare eventuali differenze nelle distribuzioni di tiro delle squadre. In alcune stagioni, potrebbero mancare i dati di alcune squadre a causa di spostamenti di franchigia o cambi di nome, come spiegato in precedenza. Tuttavia, complessivamente, per ogni stagione sono presenti i dati di trenta squadre.

Questa distribuzione è illustrata nella Figura 2.3, che mostra come, con il passare del tempo, le squadre tendano sempre più a usufruire del tiro da tre punti. È sufficiente confrontare le medie di questa statistica nelle stagioni 1996/97 e 2022/23 per osservare un aumento di oltre mille tentativi di tiri da tre punti per squadra, e allo stesso tempo, una diminuzione di circa seicento

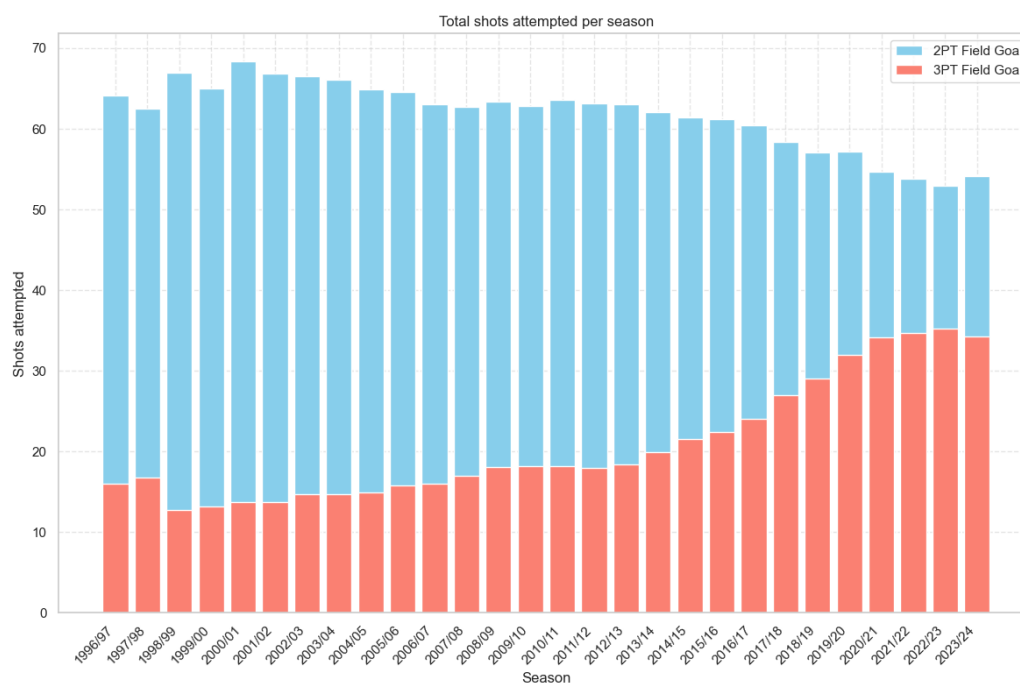


Figure 2.2: Distribuzione dei tiri tentati.

tentativi di tiri da due punti per squadra per stagione.

Un dato interessante è che nonostante il forte aumento del numero di tiri tentati, la percentuale di realizzazione rimanga stabile intorno al 38% (Figura 2.4). Questo fenomeno suggerisce che, nonostante la crescente propensione delle squadre a tentare tiri da tre punti, l'efficacia del tiro rimane costante nel tempo.

Proprio per questo motivo, si è deciso di utilizzare il numero di tiri tentati anziché quelli effettivamente realizzati come parametro di valutazione. Analizzare i tiri tentati fornisce una misura più accurata dell'attività offensiva complessiva di una squadra e della sua tendenza a sfruttare le opportunità di tiro, indipendentemente dalla percentuale di realizzazione. Inoltre, considerare i tiri tentati consente di valutare in modo più approfondito l'impatto delle strategie offensive sul punteggio complessivo, tenendo conto del volume e della frequenza dei tiri tentati.

Per quanto riguarda i tiri da due punti, è possibile osservare un significativo

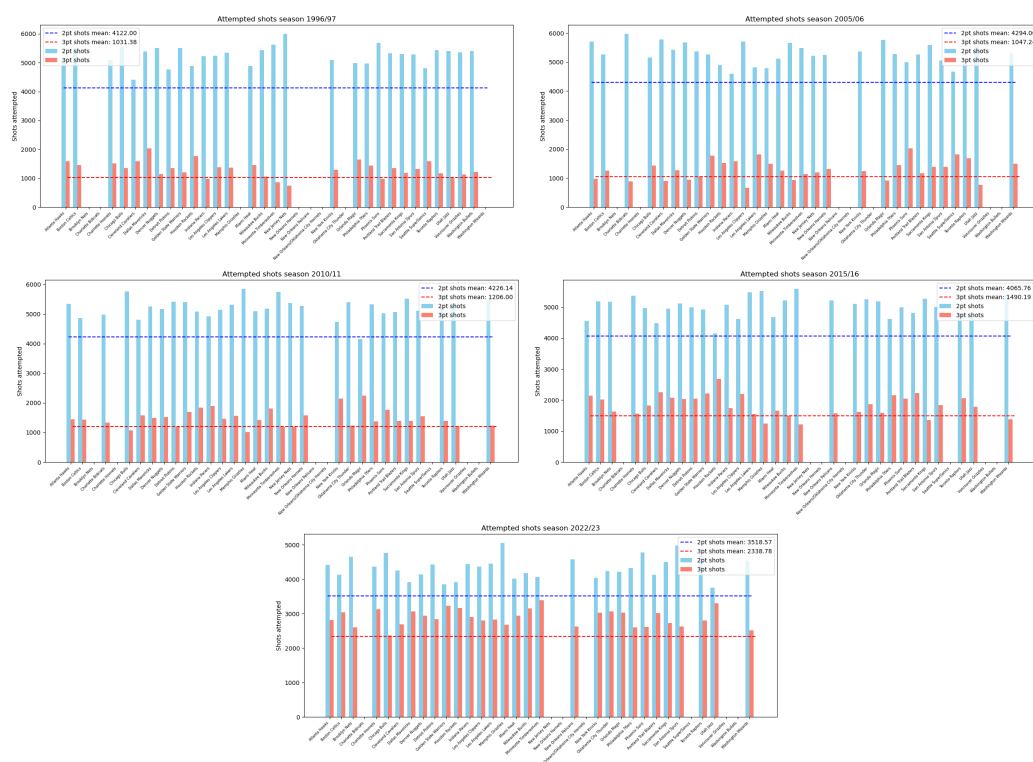


Figure 2.3: Distribuzione dei tiri da due e tre punti

calo nel numero di tentativi dal 1996 al 2024 (Figura 2.5). Tuttavia, nello stesso periodo, la percentuale di realizzazione è aumentata leggermente, passando dal 48% a quasi il 55%.

Questi due trend suggeriscono un cambiamento nelle tattiche di gioco delle squadre NBA. Tuttavia, nel caso dei tiri da due punti, la stabilità nel numero di tiri realizzati e l'aumento delle percentuali di realizzazione indicano che le squadre sono diventate progressivamente più selettive nei tiri da due punti tentati. È evidente, anche agli occhi dei meno esperti, che alcune zone del campo vengono sfruttate sempre meno dalle squadre NBA, come ad esempio il tiro dalla media distanza, ritenuto un "tiro a bassa percentuale". Molte squadre, anziché tentare un tiro da due punti lontano dal canestro, preferiscono allontanarsi ulteriormente ed effettuare un tiro da tre punti. Al contrario, è sempre più frequente vedere schiacciate o layup nei pressi del

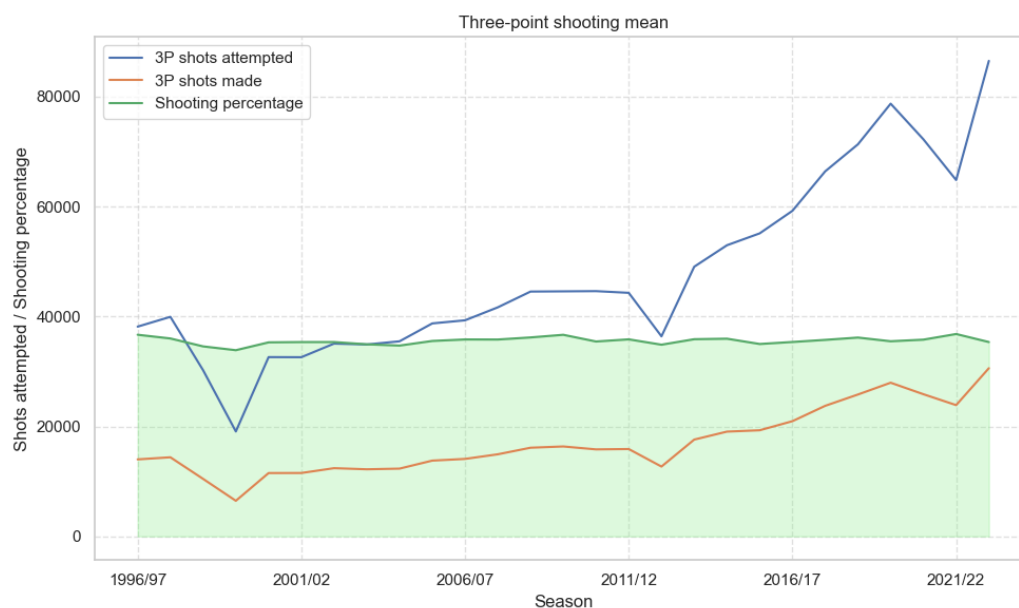


Figure 2.4: Trend tiri da tre punti tentati e realizzati con relativa media.

ferro.

Questo fenomeno può essere interpretato come una risposta strategica all'evoluzione del gioco, dove l'efficienza e la selezione del tiro sono diventate prioritarie per massimizzare il rendimento offensivo. La tendenza a favorire tiri a maggiore valore (come il tiro da tre punti) e ad alta percentuale di realizzazione (come i layup e le schiacciate) riflette un approccio sempre più analitico e strategico alla gestione delle opportunità di tiro.

2.2 Metodologie

Per comprendere se il tiro da tre punti ha influenzato in modo statistico l'NBA e se ci sia stato un incremento nel numero di vittorie di una squadra negli ultimi anni, sono state analizzate le serie storiche attraverso la rilevazione dei change point. Questa analisi è in grado di determinare se sono presenti cambiamenti nel comportamento di una serie storica e se questi coincidono con eventi specifici o con l'introduzione di nuove regole nella NBA.

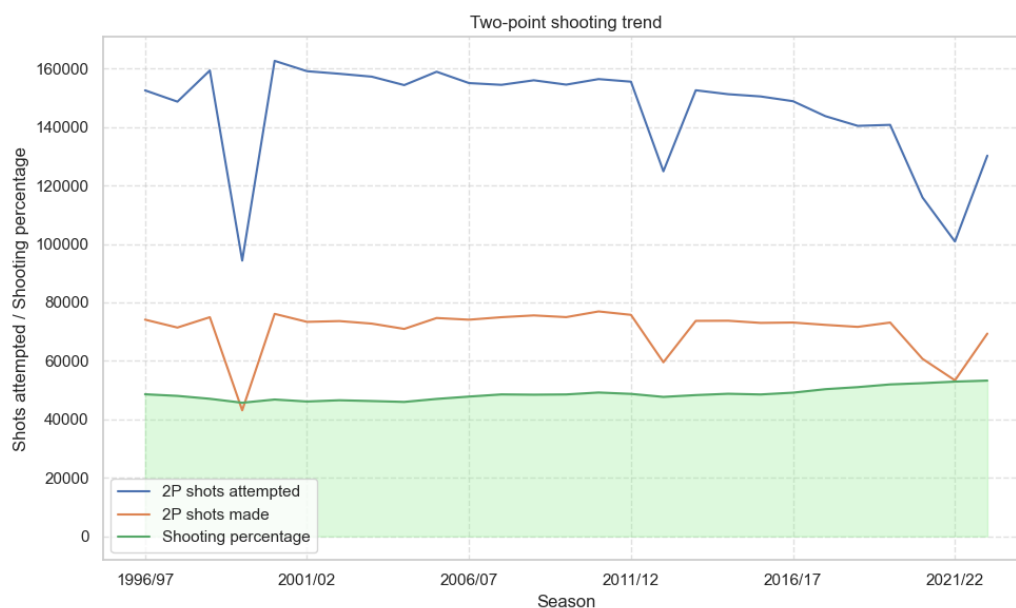


Figure 2.5: Trend tiri da due punti tentati e realizzati con relativa media.

Per stabilire se esiste una correlazione e una causalità tra i tiri da tre punti effettuati e la media punti segnati a partita, sono state utilizzate diverse tecniche statistiche. Inizialmente, è stata analizzata la relazione tra le due variabili utilizzando la correlazione di Pearson e la cross-correlation, al fine di identificare se i valori di una serie siano predittivi dei valori futuri di un'altra serie. Successivamente, è stato impiegato il test di causalità di Granger per determinare una relazione causa-effetto tra le variabili.

Per completare la ricerca, sono stati effettuati diversi test statistici. Prima di tutto, è stata determinata la presenza di tendenze positive o negative attraverso il test di Mann-Kendall sui ranghi. Successivamente, per verificare se una serie sia stazionaria o meno, sono stati utilizzati i test Augmented Dickey-Fuller (ADF) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS) basati su radici unitarie.

2.2.1 Serie temporali

Le serie temporali sono definite come dati raccolti e analizzati in ordine cronologico, con intervalli temporali equidistanti, come ad esempio una stagione sportiva. In questo studio, le serie temporali vengono utilizzate per analizzare l'andamento dei tiri da due e da tre punti nelle stagioni NBA dal 1996/97 al 2023/24. Questo metodo è scelto perché consente di osservare come il gioco è cambiato nel tempo e di identificare tendenze a lungo termine che possono essere influenzate da cambiamenti regolamentari o strategici. La variabile indipendente nei metodi delle serie temporali è unicamente il tempo. Infatti, i metodi delle serie temporali si basano sulle leggi e tecniche della scienza statistica, una branca della matematica che si occupa della raccolta, classificazione, analisi e interpretazione di dati numerici in grado di descrivere un fenomeno. Questi metodi non mirano a spiegare o individuare la struttura causale del problema, ma si concentrano sul rilevare l'andamento e l'evoluzione nel tempo[11]. Attraverso un'equazione analitica, le serie temporali consentono di riconoscere, comprendere e descrivere l'evoluzione di un fenomeno nel tempo. L'equazione analitica può essere scritta come:

$$y_t = S_t + T_t + R_t \quad (2.1)$$

ed è composta dai seguenti componenti:

- **S_t Seasonal:** rappresenta la componente stagionale e cattura i pattern regolari e ripetitivi all'interno di una serie temporale.
- **T_t Trend:** rappresenta la componente di trend, che indica la direzione generale (stazionaria, in aumento, in diminuzione) a lungo termine della serie temporale.
- **R_t Residual:** rappresenta la componente residua o irregolare, che include tutte le variazioni non spiegate dalle componenti stagionali e di trend, spesso considerate come rumore o variazioni casuali.

Nella Figura 2.6, sono mostrate le serie storiche dei tentativi di tiro effettuati da tre punti e da due punti in ogni stagione NBA dalla stagione

1996/97 alla stagione 2023/24. La figura include anche la componente di trend relativa a ciascuna serie temporale e la media dei tiri tentati nella serie temporale.

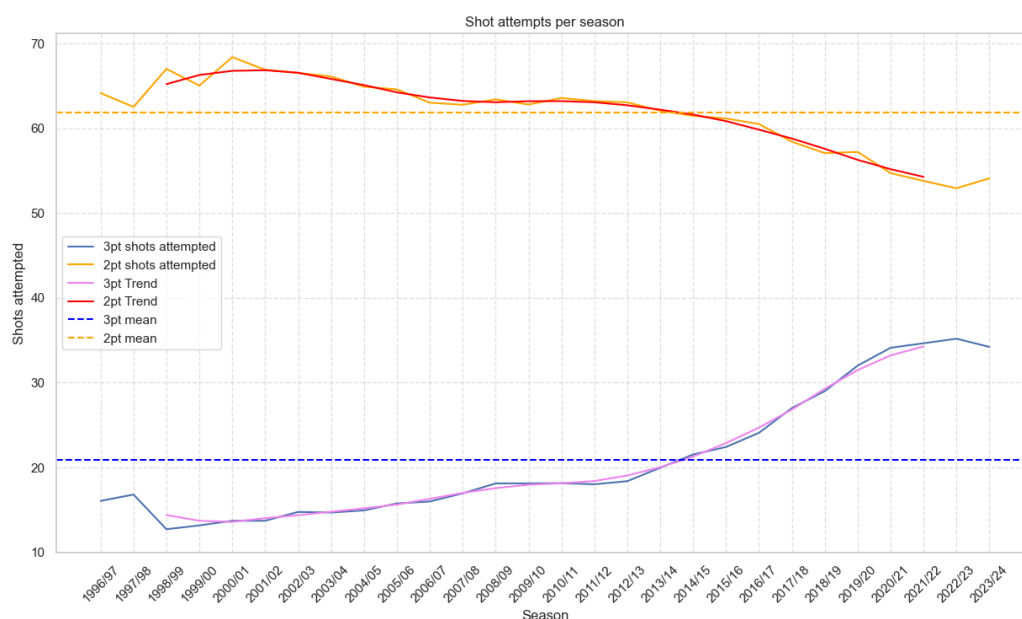


Figure 2.6: Serie temporali di tiri da due e tre punti per stagione.

La decomposizione delle serie temporali implica la suddivisione di una serie nei suoi componenti fondamentali. Questo processo è comunemente utilizzato per analizzare le serie temporali e fornire informazioni ai modelli di previsione. La decomposizione offre un modo strutturato per affrontare un problema di previsione delle serie temporali, sia in termini generali di complessità della modellazione, sia in termini specifici su come ciascun componente viene catturato in modo ottimale all'interno di un modello specifico. In particolare, la decomposizione consente di distinguere le diverse componenti di una serie temporale, come la componente stagionale, il trend e il residuo, permettendo una comprensione più approfondita del fenomeno analizzato. Questo approccio facilita la modellazione e la previsione, rendendo possibile l'identificazione delle variazioni regolari e delle tendenze a lungo termine, nonché delle fluttuazioni casuali che possono influenzare i dati.

In sintesi, la decomposizione delle serie temporali non solo aiuta a scomporre e analizzare i dati complessi, ma fornisce anche una base solida per sviluppare modelli di previsione più accurati e informati.

2.2.2 Rilevamento dei Changepoint

Dall'analisi delle componenti di una serie storica è possibile determinare un changepoint, ovvero un punto in cui i dati subiscono un cambiamento sostanziale nel loro andamento. Il metodo si basa sul calcolo della discrepanza tra le diverse parti di una serie temporale osservata. Se le due parti non presentano un cambiamento effettivo, il valore risultante della discrepanza sarà basso, in caso contrario si registrerà un picco, che permette di individuare con precisione il punto in cui avviene il cambiamento.

Il metodo utilizzato per individuare i punti di cambiamento nella serie temporale dei tiri effettuati, rappresentata in Figura 2.6, è il Bayesian Estimator of Abrupt Change (BEAST), documentato in [12]. Questo metodo utilizza la regressione bayesiana per isolare i segnali periodici e di tendenza da una serie temporale, riuscendo così a individuare cambiamenti repentini nei due segnali isolati. BEAST è in grado di estendere le analisi convenzionali delle tendenze, adattando sia le tendenze lineari che quelle non lineari, e contemporaneamente separando le tendenze dalla stagionalità.

Utilizza la media del modello bayesiano e comprende tutti i modelli candidati invece di considerarne esclusivamente uno, riuscendo a dedurre dinamiche non lineari e distinguendole dai metodi di analisi delle tendenze esistenti.

Il modello BEAST è basato su una serie temporale $D = \{t_i, y_i\}_{i=1, \dots, n}$ composta da tre componenti principali: stagionalità, tendenza e cambiamenti improvvisi, oltre al rumore.

$$y_i = S(t_i; \Theta_S) + T(t_i; \Theta_T) + \varepsilon_i \quad (2.2)$$

E' necessario assumere che il rumore, rappresentato da ε , sia gaussiano una deviazione standard σ , in modo da catturare il resto dei dati non spiegati dalla componente stagionale $S(\cdot)$ stagionale e dalla componente di tendenza

$T(\cdot)$. La pratica comune prevede di adottare modelli lineari generali per parametrizzare $S(\cdot)$ e $T(\cdot)$. I changepoint sono implicitamente codificati nei parametri Θ_S e Θ_T dei segnali di stagionalità e tendenza. Paragonandolo alla regressione lineare, il tempo t e i dati y sono rispettivamente variabili indipendenti e dipendenti; Θ_S e Θ_T sono parametri che vengono stimati dai dati D . In particolare, il segnale stagionale $S(t)$ viene approssimato come un modello armonico a tratti, definito rispetto a p nodi nel tempo in $\xi_k = 1, \dots, p$. I nodi, vengo utilizzati per dividere la serie temporale in intervalli $(p + 1)$ $[\xi_k, \xi_{k+1}]$, $k = 0, \dots, p$

$$S(t) = \sum_{l=1}^{L_k} \left[a_{k,l} \cdot \sin\left(\frac{2\pi lt}{P}\right) + b_{k,l} \cdot \cos\left(\frac{2\pi lt}{P}\right) \right]; \xi_k \leq t \leq \xi_{k+1}, k = 0, \dots, p \quad (2.3)$$

In questo caso P rappresenta il periodo del segnale stagionale(ad esempio, una stagione sportiva) mentre L_k indica l'ordine armonico del k -esimo segmento ed è un parametro specifico del segmento sconosciuto; i coefficienti $\{a_{k,l}\}_{l=1, \dots, L_k}$ indicano i parametri per seni e coseni. L'andamento $T(t)$ viene modellato come una funzione lineare a tratti rispetto m nodi in $\tau_j = 1, \dots, m$, che sono utilizzati per dividere l'arco temporale in $(m+1)$ intervalli $[\tau_j, \tau_{j+1}]$, $j = 0, \dots, m$.

$$T(t) = a_j + b_j t \quad (\tau \leq t < \tau_{j+1}, j = 0, \dots, m) \quad (2.4)$$

con a_j e b_j indichiamo i coefficienti di funzioni lineari. Il numero di changepoints e il loro momento temporale corrispondente sono parametri sconosciuti di τ_j . I parametri correlati vengono riclassificati in due gruppi differenti: uno dedicato ai parametri della struttura del modello M e uno ai parametri dei coefficienti specifici del segmento β_M :

$$\begin{aligned} M &= \{m\} \cup \{\tau_j\}_{j=1, \dots, m} \cup \{p\} \cup \{\xi_k\}_{k=1, \dots, p} \cup \{L_k\}_{k=0, \dots, p} \\ \beta_M &= \{a, b_j\}_{j=0, \dots, m} \cup \{a_{k,l}, b_{k,l}\}_{k=0, \dots, p; l=1, \dots, L_k} \end{aligned} \quad (2.5)$$

La struttura del modello M include il numero e il tempo che corrispondono ai changepoint stagionali e di tendenza, e l'ordine armonico stagionale. Con β_M

rapresentiamo i coefficienti del modello sezionato di spostamento stagionale e tendenziale. Andiamo a considerare l'equazione 2:

$$y(t_i) = x_m(t_i) \beta_m + \varepsilon_i \quad (2.6)$$

Trovare la struttura adatta del modello M per il nostro problema è analogo alla selezione del miglior sottoinsieme di variabili in una regressione lineare semplice. Una volta individuata la struttura ottimale, i coefficienti B risultano semplici da stimare. Questo problema viene affrontato utilizzando l'inferenza bayesiana, che consente di ottenere i valori ottimali di questi parametri e le loro distribuzioni di probabilità a posteriori

$p(\beta_M, \sigma^2, M \{Y_t\}_{t=1, \dots, n})$. Siccome non è trattabile dal punto di vista analitico, il campionamento MCMC (Markov Chain Monte Carlo) viene utilizzato per generare una realizzazione di campioni casuali per l'inferenza a posteriori. L'algoritmo MCMC impiegato è un campionario ibrido che incorpora un campionamento MCMC con salto reversibile. Questo algoritmo genera una catena di campioni posteriori di lunghezza

$N \{ \beta_M^{(i)}, \sigma^{2(i)}, v^{(i)}, M^{(i)} \}_{i=1, \dots, n}$ (v è un iperparametro di dispersione extra, introdotto per riflettere ulteriormente la vaga conoscenza della grandezza dei coefficienti del modello β_M). la catena è in grado di catturare tutte le informazioni essenziali per dedurre le dinamiche del terreno, incluse tendenze, variazioni stagionali e cambiamenti improvvisi. La combinazione delle singole stime riesce a fornire la media finale del modello Bayes (Bayesian Model Averaging, BMA): $\hat{y}(t) \approx \sum_{i=1}^N x_M^{(i)}(t) \beta_M^{(i)} / N$. Dalla struttura del modello campionato M è possibile stimare il numero totale medio di changepoints di stagionalità e tendenza, oltre che i momenti esatti o l'intervallo di tempo in cui sono avvenuti i cambiamenti di stagionalità o tendenza per ciascun modello tra quelli campionati. Allo stesso modo, stimando un changepoint, si è in grado di ricavare l'intervallo possibile in cui tale cambiamento potrebbe essere avvenuto. Alla stessa maniera da un changepoint dato si è in grado di ricavare l'intervallo credibile.

Le informazioni ricavate da BEAST non si limitano alla stagionalità, alle tendenze o ai changepoint di queste ultime (*scp* o *tcp*), né agli ordini armonici

dei singoli segmenti stagionali. BEAST fornisce anche delle statistiche diagnostiche, come la probabilità di osservare un *scp* o un *tcp* in un determinato momento, la distribuzione di probabilità del numero totale di *scp* o *tcp* e la probabilità di avere un trend positivo o negativo.

L'applicazione di questo algoritmo in questo studio è stata fatta presupponendo l'assenza di stagionalità nei dati, quindi verrà tenuta in considerazione esclusivamente la presenza di trend.

2.2.3 Test di Mann-Kendall

Questo test è applicato per determinare se ci sono tendenze significative nel numero di tiri tentati da due e da tre punti nelle diverse stagioni. Utilizzando il test di Mann-Kendall[13], possiamo verificare se le osservazioni mostrano un trend crescente o decrescente, aiutandoci a comprendere meglio l'evoluzione del gioco nel tempo. Il test di Mann-Kendall viene utilizzato per analizzare i dati delle serie temporali, essendo in grado di individuare tendenze monotone, ovvero tendenze costantemente crescenti o decrescenti. Essendo non parametrico, risulta particolarmente adatto per le distribuzioni dove i dati potrebbero non seguire una distribuzione normale.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad \text{sgn}(z) = \begin{cases} 1 & \text{se } z > 0 \\ 0 & \text{se } z = 0 \\ -1 & \text{se } z < 0 \end{cases} \quad (2.7)$$

dove:

- n : indica la lunghezza della serie.
- x_i e x_j : indicano le osservazioni a tempo i e a tempo j , rispettivamente dove $i < j$.

Con varianza data da:

$$\text{var}(S) = \frac{n(n-1)(2n+5)}{18} \quad (2.8)$$

La significatività di un trend è testata con la statistica test standardizzata Z :

$$Z = \frac{S}{\sqrt{Var(S)}} \quad (2.9)$$

Se i dati risultano avere una correlazione seriale, potrebbero influenzare a livello significativo una serie temporale (p value). Si potrebbe cascare in interpretazione errate, per superare questo problema, in letteratura esistono differenti varianti di Mann-Kendall, e per questa ricerca è stato selezionato quello proposto da Hamed e Rao[14]. Quindi, l'ipotesi di test e le due ipotesi sono le seguenti:

- **Ipotesi nulla H_0** : non è presente alcun trend nei dati.
- **Ipotesi alternativa H_1** : è presente un trend nei dati.

2.2.4 Test della stazionarietà

La stazionarietà in una serie temporale è caratterizzata da proprietà statistiche costanti e dall'assenza di stagionalità. Una serie temporale viene considerata stazionaria quando sia la media che la deviazione standard tendono a rimanere costanti nel tempo e non sono riconoscibili trend o stagionalità nei dati. Al contrario, una serie temporale non stazionaria tende a variare le sue proprietà statistiche nel tempo e solitamente presenta componenti stagionali e di trend. In questi casi, la media, la deviazione standard e i modelli presenti nella serie temporale variano, rendendo necessario considerare queste caratteristiche dinamiche nelle analisi.

Alcuni test appartengono alla categoria dei test della radice unitaria, che risultano essere il metodo migliore per valutare la stazionarietà di una serie temporale. La presenza di una radice unitaria implica che la serie temporale mostri una tendenza stocastica o casuale nel tempo. Per verificare la presenza di una radice unitaria in una serie temporale, si utilizza l'equazione:

$$Y_t = \alpha Y_{t-1} + X_e \quad (2.10)$$

Se il valore di α è uguale a uno, allora è presente una radice unitaria nella serie. In questa equazione, Y_t rappresenta il valore della serie temporale al tempo t e X_e è una variabile esogena, che anch'essa è una serie temporale. La presenza di una radice unitaria indica che la serie temporale è non stazionaria. Per valutare questo aspetto, vengono comunemente utilizzati due test:

- Augmented Dickey-Fuller Test (ADF)
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Augmented Dickey Fuller test(ADF)

Il test ADF (Augmented Dickey-Fuller) [15] è un test statistico essenziale utilizzato per determinare se una serie temporale è stazionaria o non stazionaria, verificando la presenza di una radice unitaria. La stazionarietà è cruciale in analisi temporali poiché implica che le proprietà statistiche della serie, come la media e la varianza, non variano nel tempo. Nel contesto di questa tesi, il test ADF è fondamentale per comprendere se il numero di tiri tentati in una partita di basket NBA è influenzato da tendenze a lungo termine o se le fluttuazioni sono casuali. Determinare la stazionarietà delle serie temporali tramite il test ADF è essenziale per stabilire se e come le serie debbano essere trattate, ad esempio mediante differenziazione, per migliorare la loro adattabilità alle analisi predittive. Il test sviluppato da Dickey e Fuller si concentra sulla radice unitaria, verificando l'ipotesi nulla che $\alpha = 1$ nell'equazione:

$$y_t = \alpha + \beta t + \gamma y_{t-1} + \phi \Delta y_{t-1} + \epsilon_t \quad (2.11)$$

dove:

- Δy_{t-1} indica la prima differenza della serie al tempo $t - 1$.
- γy_{t-1} indica il ritardo di uno della serie temporale.
- ϵ_t indica il termine di errore.

Alla base del test c'è l'intuizione che, se una serie temporale è caratterizzata da un processo di radice unitaria, il ritardo nella serie non sarà in grado di prevedere informazioni rilevanti riguardo al cambiamento oltre a quelle ottenute dai cambiamenti ritardati. Pertanto, Dickey e Fuller, per verificare l'autocorrelazione di ordine superiore, aumentano l'equazione con ritardi di ordine superiore, ottenendo il test Augmented Dickey-Fuller (ADF) per le serie temporali.

L'ipotesi nulla del test ADF è che esiste una radice unitaria in un modello autoregressivo (AR), il che implica che i dati non sono stazionari. L'ipotesi alternativa è generalmente la stazionarietà, anche se può variare in base alla versione del test utilizzata.

Kwiatkowski-Phillips-Schmidt-Shin Test (KPSS)

Il test KPSS è stato sviluppato per integrare i test di radice unitaria, poiché questi ultimi sono considerati meno efficaci nel rilevare processi con trend prossimi alla radice unitaria e di lungo termine.

Nel contesto di questo studio, il test KPSS è utilizzato congiuntamente al test ADF per confermare l'analisi della stazionarietà della serie temporale. Mentre il test ADF verifica la presenza di una radice unitaria, il test KPSS fornisce una prospettiva aggiuntiva che arricchisce la comprensione della struttura dei dati. Questo test determina se una serie temporale è stazionaria attorno a un trend deterministico.

L'ipotesi nulla nel test KPSS stabilisce che la serie temporale sia stazionaria attorno a un trend deterministico. Pertanto, a differenza del test ADF, un risultato che supporta l'ipotesi nulla indica la stazionarietà della serie.

Entrambi i test sono cruciali per valutare la stazionarietà di una serie temporale, e pertanto vengono utilizzati congiuntamente per ottenere una visione più completa delle proprietà della serie temporale nella mia ricerca.

2.2.5 Analisi della Correlazione

In statistica, per correlazione si intende una relazione tra due variabili tale per cui ad ogni valore della prima corrisponde un valore della seconda, secondo una certa regolarità. Individuare una correlazione può, in alcuni casi, rivelarsi vantaggioso poiché una relazione predittiva può essere utilizzata nella pratica. La correlazione viene definita da un coefficiente, indicato con ρ , il cui valore può variare tra -1 e +1. La significatività statistica di una correlazione viene indicata con il p-value, che rappresenta una misura di probabilità utilizzata nella verifica delle ipotesi.

La correlazione ha le seguenti caratteristiche:

- ρ positivo: indica una correlazione positiva, quindi al variare di una variabile, un'altra variabile aumenta nella stessa direzione.
- ρ vicino a zero: indica assenza di correlazione, quindi due variabili non sono correlate tra loro.
- ρ negativo: indica una correlazione negativa, quindi al variare di una variabile, un'altra variabile varia in modo inverso.

Uno dei coefficienti maggiormente utilizzati in statistica è la correlazione di Pearson, che viene definita come:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.12)$$

In questo contesto deve essere considerato lo studio delle correlazioni incrociate, in quanto sono una generalizzazione della misura di correlazione, poiché considerano il ritardo di un segnale rispetto ad un altro. Pertanto, la correlazione incrociata risulta essere fondamentale nella valutazione della relazione causale tra due serie temporali, tenendo traccia della loro somiglianza.

Questo valore misura la somiglianza tra due serie temporali in funzione del lasso di tempo, permettendo la scoperta di modelli e dipendenze. Quando si considera il ritardo 0, i risultati ottenuti sono i coefficienti di Pearson. La correlazione incrociata è utile perché verifica se i valori di una serie temporale

sono predittivi dei valori futuri di un'altra serie temporale, permettendo di capire se una serie temporale è un indicatore anticipatore per un'altra. Anche in questo caso l'intervallo che può assumere il coefficiente di correlazione delle serie storiche varia tra -1 e $+1$, mantenendo le stesse caratteristiche elencate precedentemente per la correlazione.

Non basta la presenza di una correlazione per dedurre una relazione causale. Per determinare una relazione causale tra variabili, in entrambe le direzioni, la correlazione da sola non è sufficiente. In questi casi, esiste una definizione di causalità nelle serie temporali che permette di determinare il senso causa-effetto basandosi sul principio che la prima serie aiuta a prevedere la seconda. Il primo concetto di causalità per i dati di serie temporali è stato suggerito da Granger[16].

2.2.6 Test di causalità di Granger

Nel contesto di questo studio, il test di causalità di Granger riveste un'importanza fondamentale poiché fornisce un metodo per valutare l'influenza dei tiri tentati da due e tre punti sul numero totale di punti segnati durante le partite NBA nel corso del tempo. Questo test è essenziale per analizzare le possibili relazioni tra queste variabili e per prevedere le tendenze nel punteggio delle partite.

Il test di causalità di Granger si basa sulla capacità di dimostrare, attraverso una serie di test t e F su valori ritardati delle variabili X (tiri tentati da due e tre punti) e Y (numero di punti segnati), che le informazioni storiche di X sono utili per prevedere Y in modo significativo nel tempo. Se tale relazione può essere stabilita, si può affermare che X Granger-causa Y , indicando che le variazioni nei tiri tentati possono predire in qualche modo le variazioni nel punteggio totale delle partite.

Questo approccio richiede che i dati siano stazionari, il che può essere verificato con test appropriati come quelli menzionati precedentemente. In caso contrario, è possibile rendere stazionari i dati mediante tecniche come la differenziazione.

Il test di Granger si basa su modelli autoregressivi vettoriali (VAR) delle serie temporali endogene, come descritto da Granger e Lütkepohl (2005)[17]. Tale modello consente di considerare le dinamiche reciproche tra più variabili nel tempo, rendendo il test adatto per esaminare le relazioni complesse tra tiri tentati e punteggio nella NBA.

Verranno addestrati due modelli VAR:

- Modello ristretto: presuppone che Y sia linearmente dipendente solo dai valori passati di se stessa, con coefficienti lineari γ_i e un termine di rumore dipendente dal tempo e_t :

$$Y_t = \gamma + \sum_{i=1}^p \gamma_i Y_{t-i} + e_t \quad (2.13)$$

- Modello senza restrizioni: presuppone che Y sia linearmente dipendente dai valori passati di X e Y , determinati dai coefficienti α_i , β_i e da un termine di rumore dipendente dal tempo u_t :

$$Y_t = \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{i=1}^p \beta_i X_{t-i} + u_t \quad (2.14)$$

L'ipotesi nulla non formalizzata è che il secondo modello non aggiunga informazioni o non fornisca un miglioramento significativo rispetto al primo modello. Tuttavia, deve essere formalizzata un'ipotesi verificabile:

- Ipotesi nulla H_0 : la serie temporale X non causa Y secondo Granger.
- Ipotesi alternativa H_1 : la serie temporale X causa Y secondo Granger.

Per rifiutare l'ipotesi nulla, dobbiamo ottenere un p-value che rispetti un certo livello di significatività (ad esempio, $p = 0,005$), in modo da concludere che ci sono prove sufficienti per affermare che la serie temporale X causa la serie temporale Y secondo Granger.

2.2.7 Regressione Lineare

In questa analisi, la regressione lineare rappresenta uno strumento statistico utilizzato per esplorare la relazione tra la percentuale di vittorie delle squadre NBA (variabile dipendente) e diverse variabili indipendenti continue. Queste includono i tentativi di tiro da due e da tre punti, il ritmo di gioco (Pace), i rating offensivo e difensivo, e i punti segnati per partita.

Il modello di regressione lineare è espresso dalla seguente equazione:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Dove:

- \hat{y} : indica la variabile dipendente.
- x : indica la variabile indipendente.
- β_0 : è l'intercetta della retta di regressione.
- β_1 : è la pendenza della retta di regressione.

L'obiettivo di questo modello è trovare i valori ottimali dei coefficienti $\beta_0, \beta_1, \dots, \beta_p$ in modo che la retta di regressione ottenuta rappresenti al meglio la relazione tra le variabili, minimizzando l'errore tra i valori predetti (\hat{y}) e i valori osservati della variabile dipendente (y).

Questo approccio permette di analizzare in modo sistematico e quantitativo come le variabili selezionate influenzino la performance delle squadre NBA in termini di percentuale di vittorie. La regressione lineare fornisce un quadro chiaro e strutturato per esaminare l'importanza relativa di ciascuna variabile indipendente nel determinare il successo delle squadre nel campionato.

Chapter 3

Risultati

Come discusso nel capitolo precedente, l'analisi delle serie storiche consente di individuare eventuali punti di cambiamento di tendenza o stagionalità. In questo capitolo verranno presentati i risultati ottenuti dall'utilizzo dell'algoritmo BEAST per identificare i punti in cui avviene un cambiamento di tendenza. Per entrambe le serie relative ai tiri tentati, sono stati individuati due changepoint: per i tiri da tre punti nel 2000 e nel 2010, e per i tiri da due punti nel 2000 e nel 2008.

Per valutare eventuali trend, è stato applicato il test di Mann-Kendall, che ha evidenziato un trend positivo per i tiri da tre punti e un trend negativo per i tiri da due punti.

Successivamente, sono stati condotti due test sulle radici unitarie: il test Augmented Dickey-Fuller (ADF) e il test Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Entrambi i test hanno indicato la presenza di una radice unitaria nelle serie, confermando la loro non stazionarietà e l'esistenza di trend.

Lo studio si è concentrato sull'influenza dell'aumento del numero di tiri da tre punti tentati sulla media di punti segnati durante una partita NBA, confrontandola con l'influenza del numero di tiri da due punti tentati sulla stessa media di punti per partita.

Inoltre, sono state esaminate la correlazione e la causalità tra i tiri da due e tre punti tentati e la media di punti per partita in ogni stagione NBA,

utilizzando la correlazione incrociata e il test di causalità di Granger. È stato sottolineato l'importanza di verificare la stazionarietà dei dati prima di applicare il test di causalità di Granger, poiché quest'ultimo richiede serie temporali stazionarie per inferenze valide.

L'importanza dell'analisi della correlazione incrociata su serie stazionarie è enfatizzata nel libro di Derrick e Timothy [18]. Le limitazioni e i recenti sviluppi riguardo all'approccio della causalità di Granger sono discussi nell'articolo di Shojaie et al. [19], dove viene sottolineata la necessità di serie temporali stazionarie per un'adeguata inferenza sulla causalità.

Infine, utilizzando un modello di regressione lineare, è stato esaminato come i fattori analizzati influenzano la percentuale di vittorie di una squadra NBA. Il modello ha dimostrato una buona precisione nella predizione della percentuale di vittorie, indicando che questa può essere spiegata dalle variabili indipendenti utilizzate, in particolare dai tentativi di tiro da tre punti e dall'efficienza offensiva e difensiva delle squadre.

3.1 Rilevamento dei changepoint per i tiri effettuati in NBA

La Figura 3.1 mostra l'applicazione dell'algoritmo BEAST al numero di tiri da tre punti tentati dalla stagione 1996/97 alla stagione 2023/24.

Nel grafico vengono mostrati:

- **Trend:** Viene mostrato l'andamento dei tiri da tre punti tentati nel tempo, con linee nere verticali che indicano i changepoint rilevati dall'algoritmo BEAST.
- **Pr(tcp):** Mostra la probabilità di occorrenza dei changepoint nel tempo.
- **Slpsgn:** Indica la probabilità che la pendenza sia positiva (parte rossa), zero (parte verde) o negativa (parte blu).

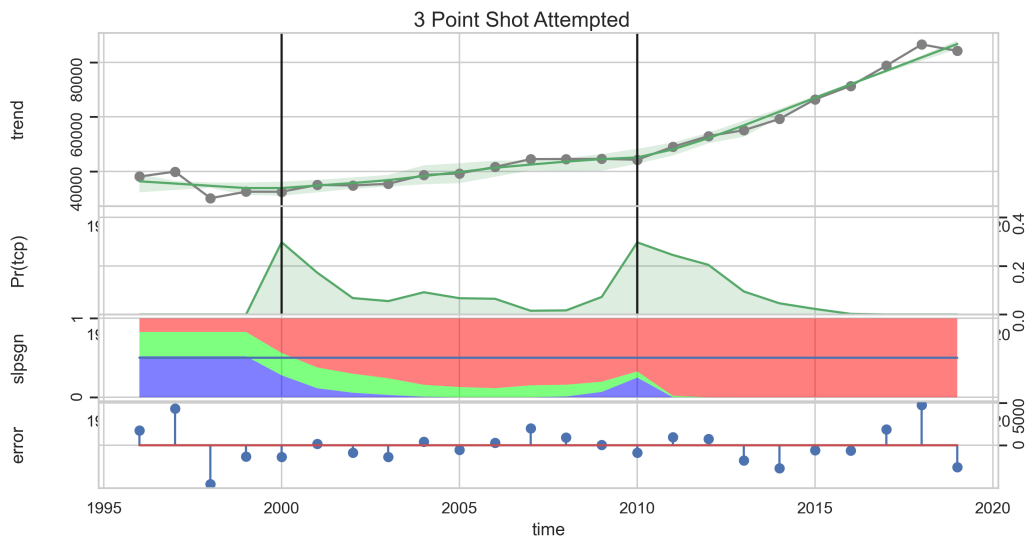


Figure 3.1: Rilevamento dei changepoint nel numero di tiri da tre punti effettuati dal 1996

- **Error:** Visualizza l'errore associato ai dati.

Oltre alla rilevazione dei changepoint, BEAST fornisce anche delle statistiche aggiuntive che indicano la probabilità che un changepoint si verifichi in una determinata stagione. Nella Tabella 3.1 è riportata la probabilità che un changepoint sia presente in una specifica stagione:

tcp#	time (cp)	prob(cpPr)
1	2010	0.84150
2	2000	0.59283

Table 3.1: Probabilità di un changepoint tra il 1996 e 2024

Le probabilità riportate in tabella sono il risultato dell'algoritmo BEAST, che permette di determinare in quanti modelli MCMC simulati è presente un changepoint nel momento t .

La deviazione standard (Figura 3.2) intorno alla media dei dati ha un ruolo fondamentale per comprendere come la distribuzione dei dati influisca sulla scelta dei changepoint.

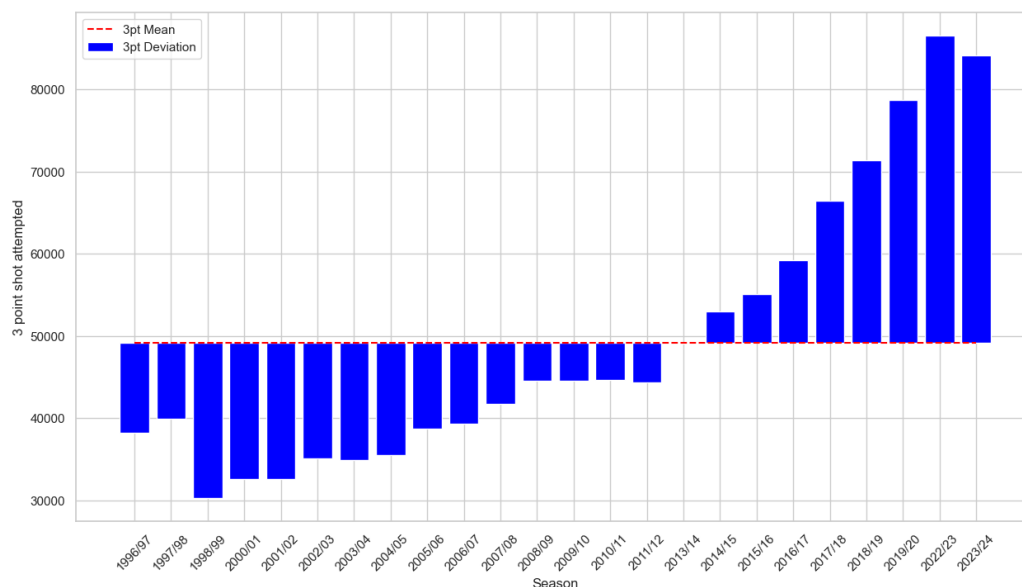


Figure 3.2: Deviazione standard e media dei tiri da 3 punti tentati.

Dal grafico in Figura 3.1 si hanno due changepoint:

- Il primo changepoint può essere attribuito all'introduzione di una nuova regola, che permetteva alle difese NBA di utilizzare la *difesa a zona*. Questo tipo di difesa, molto comune nel basket europeo e collegiale, differisce dalla classica difesa a uomo in quanto ogni difensore non si occupa di un singolo avversario, ma di una zona del campo. Prima del 2001, nella NBA era vietato l'utilizzo della difesa a zona poiché un giocatore non poteva posizionarsi in maniera intermedia e prolungata tra l'attaccante diretto e la palla senza marcare nessuno dei due. Questo regolamento rendeva impossibile l'adozione della difesa a zona. Nel 2001, con l'approvazione di un nuovo regolamento da parte della lega[20], gli allenatori hanno iniziato a integrare la difesa a zona nelle tattiche di gioco delle squadre. Questa modifica ha portato a una maggiore congestione dell'area intorno al canestro, rendendo più difficile segnare punti nelle vicinanze del ferro. Di conseguenza, le squadre NBA hanno dovuto migliorare la spaziatura sul campo e cercare più tiri aperti da tre punti, che risultano essere concessi più facilmente dalla

difesa a zona.

- Il secondo changepoint è legato all'esplosione del fenomeno Stephen Curry e dei suoi Golden State Warriors. Curry è stato scelto dai Golden State Warriors nel draft del 2009 e, sotto la guida di Don Nelson, ha iniziato a muovere i primi passi nella NBA. Già nella prima stagione, il suo potenziale è emerso chiaramente: ha concluso il campionato con 17,5 punti, 4,5 rimbalzi e 5,9 assist a partita, ottenendo il secondo posto nella classifica del "*Rookie Of The Year*". Curry ha mostrato fin dall'inizio un modo di interpretare il gioco molto diverso rispetto a quello tradizionale. La maggior parte dei suoi tiri da tre punti venivano tentati dal palleggio, piuttosto che aspettando uno scarico da un compagno, come facevano molti altri tiratori. Nel 2011, i Golden State Warriors hanno scelto al draft Klay Thompson, un altro tiratore destinato a rivoluzionare il modo di giocare. Insieme, Curry e Thompson hanno creato un duo che ha fatto del tiro da tre punti la loro arma principale. Thompson ha uno stile di gioco completamente opposto rispetto a Curry. Klay è un eccellente tiratore sugli scarichi, mentre Curry è abilissimo nel creare i propri tiri dal palleggio. Combinando queste due caratteristiche, i Golden State Warriors sono riusciti a vincere quattro titoli NBA dal 2014 al 2022, con al centro del loro sistema offensivo i cosiddetti "*Splash Brothers*". Questo successo ha influenzato enormemente tutta la lega, spingendo altre squadre ad adottare un approccio simile e aumentando notevolmente il numero di tiri da tre punti tentati in generale.

Applicando l'algoritmo BEAST all'andamento del numero di tiri da due punti tentati in stagione, si ottiene il risultato mostrato in Figura3.3.

Nella Tabella3.2 si può visualizzare la probabilità con la quale i changepoint trovati dall'algoritmo BEAST si verificano durante le stagioni indicate:

Come riportato nella Figura3.3 sono stati individuati due changepoint:

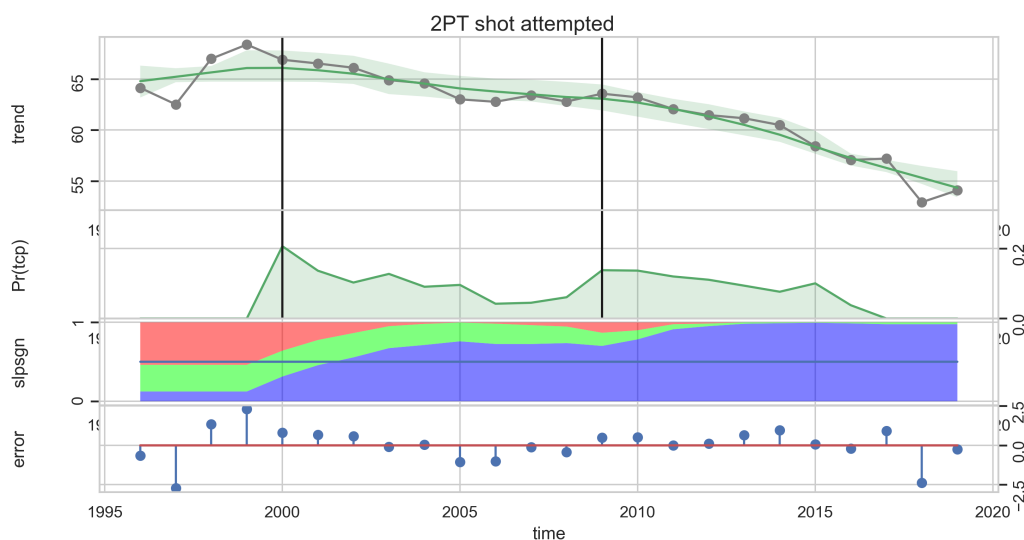


Figure 3.3: Rilevamento dei changepoint nel numero di tiri da tre punti effettuati dal 1996.

tcp#	time (cp)	prob(cpPr)
1	2000	0.79938
2	2012	0.54596

Table 3.2: Probabilità di un changepoint tra il 1996 e 2024

- Il primo è individuato nella stessa stagione del primo changepoint rilevato nella serie temporale dei tiri da tre punti. È quindi probabile che questo changepoint sia influenzato anch'esso dall'introduzione della nuova regola NBA che permette l'utilizzo della difesa a zona. Questa regola rende più difficile segnare canestri nei pressi del ferro e favorisce tiri da oltre l'arco dei tre punti.
- Il secondo changepoint trovato in questa serie temporale può essere attribuito a Daryl Morey, ex general manager degli Houston Rockets. Morey, ispirandosi al film *Moneyball*, ha creato il sistema offensivo *Moreyball* [21], applicando principi di analisi avanzata e statistiche per massimizzare i punti per possesso di una squadra NBA. Questo sis-

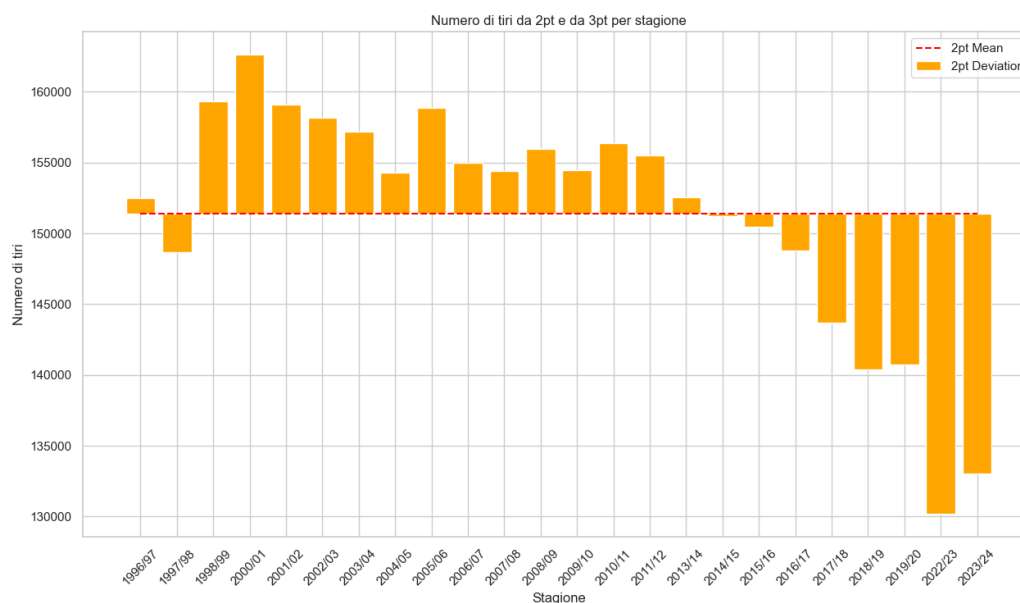


Figure 3.4: Deviazione standard e media dei tiri da 2 punti tentati.

tema si basa su una semplice regola: i tiri devono essere effettuati da posizioni con il più alto valore atteso. Per realizzare questo sistema, Morey scoraggia i giocatori dal prendere tiri dalla media distanza, in quanto considerati meno efficienti rispetto ai tiri da tre punti o ai tiri vicino al canestro. Per facilitare questa tattica, è cruciale avere giocatori in grado di offrire una buona spaziatura in campo. Questo implica avere giocatori che possono tirare da tre punti in modo da allargare le difese avversarie e creare spazi per attaccare il ferro. A partire dalla metà degli anni 2010, gli Houston Rockets hanno iniziato ad adottare questo sistema offensivo, arrivando a una sola partita dalle finali NBA con James Harden come principale giocatore della squadra. Il *Morey-ball* è stato successivamente adottato da un numero crescente di squadre NBA, portando a un cambiamento sistematico nel modo in cui viene giocato il gioco.

3.2 Rilevamento di Trend nei tiri tentati

In questa ricerca, è stato applicato il test di Mann-Kendall per verificare l'effettiva presenza di trend all'interno della serie temporale dei tiri da tre punti tentati nella NBA. L'analisi tramite l'algoritmo BEAST ha rilevato due changepoint significativi nel numero di tiri da tre punti tentati: il primo intorno al 2000 e il secondo intorno al 2010, suggerendo possibili variazioni nel trend di tiri tentati da tre punti.

Per confermare questi trend e valutare l'andamento nei segmenti identificati, è stato applicato il test Mann-Kendall ai seguenti periodi, escludendo gli anni dei changepoint per evitare distorsioni nei risultati:

- Periodo 1: 1996/97 - 1999/00
- Periodo 2: 2000/01 - 2009/10
- Periodo 3: 2010/11 - 2023/24

Ogni segmento è stato analizzato separatamente per comprendere al meglio le dinamiche di ciascun periodo e verificare statisticamente se la tendenza positiva è consistente in tutto l'intervallo di studio.

Periodo	Valore Z	Valore P	Punteggio MK	Decisione
1996-2000	-0.33968	0.73409	-2	Nessun trend
2001-2010	3.44048	0.00058	34	Trend positivo
2011-2024	3.89249	9.9218e-05	51	Trend positivo

Table 3.3: Risultati del test Mann-Kendall sui tiri da tre punti tentati.

Il secondo e il terzo periodo analizzati ottengono un p -value inferiore alla soglia di significatività, mentre il primo lo supera ampiamente. Questo risultato è dovuto principalmente a causa della scarsità di dati, che rende difficile valutare una tendenza chiara. Nel secondo e terzo periodo, invece, l'ipotesi di assenza di trend nei dati viene rigettata, confermando la presenza di trend positivi e supportando l'esistenza dei changepoint nel numero di tiri da tre

punti tentati dai giocatori NBA.

Inoltre, il test di Advanced Dickey-Fuller (ADF) è stato applicato per verificare la stazionarietà della serie temporale dei tiri da tre punti tentati (3PTA).

I risultati ottenuti sono i seguenti:

ADF Statistic	1.662877
p-value	0.998034
1% Crit. Value	-3.753
5% Crit. Value	-2.998
10% Crit. Value	-2.639

Table 3.4: Risultati del test ADF per i tiri da tre punti tentati.

Dato che l'ADF statistic è superiore ai valori critici e il p-value è molto alto, non possiamo rifiutare l'ipotesi nulla di non stazionarietà. Questo indica che la serie temporale dei tiri da tre punti tentati è non stazionaria, il che è coerente con i risultati dell'algorithm BEAST che ha identificato due changepoints significativi. La non stazionarietà può essere dovuta a questi cambiamenti strutturali nel trend dei tiri da tre punti tentati.

Per confermare ulteriormente questi risultati, è stato applicato anche il test KPSS (Kwiatkowski-Phillips-Schmidt-Shin) alla stessa serie temporale. I risultati ottenuti sono i seguenti:

KPSS Statistic	0.620649
p-value	0.020759
10% Crit. Value	0.347
5% Crit. Value	0.463
2.5% Crit. Value	0.574
1% Crit. Value	0.739

Table 3.5: Risultati del test KPSS per i tiri da tre punti tentati.

Dato che il KPSS statistic è superiore ai valori critici al 10%, 5%, e 2.5%, e il p-value è inferiore ai livelli di significatività, possiamo rifiutare

l'ipotesi nulla di stazionarietà. Questo indica che la serie temporale dei tiri da tre punti tentati è non stazionaria secondo il test KPSS, confermando ulteriormente i risultati ottenuti con il test ADF e l'algoritmo BEAST.

Gli stessi test sono stati svolti per il numero di tiri da due punti tentati. Anche in questo caso, i periodi sono stati divisi intorno ai due changepoint trovati dall'algoritmo BEAST, come nel caso dei tiri da tre punti, in modo da applicare il test di Mann-Kendall evitando distorsioni nei risultati:

- Periodo 1: 1996/97 - 1999/2000.
- Periodo 2: 2001/02 - 2011/12.
- Periodo 3: 2013/14 - 2022/2023.

Periodo	Valore Z	Valore P	Punteggio MK	Decisione
1996-2000	0.33968	0.73409	2	Nessun trend
2001-2012	-2.64689	0.00812	-35	Trend negativo
2013-2023	-4.04819	5.16142e-05	-53	Trend negativo

Table 3.6: Risultati del test Mann-Kendall sui tiri da due punti tentati.

I risultati ottenuti da questo test evidenziano un cambiamento significativo nel numero di tiri da due punti tentati, con una chiara tendenza negativa a partire dalla stagione 2000/01. Nel primo segmento non è stato rilevato nessun trend: il valore Z è molto vicino a zero e il p -value è molto alto, suggerendo che non può essere rifiutata l'ipotesi nulla di assenza di trend. Tuttavia, è doveroso tenere in considerazione la breve durata del periodo, che potrebbe aver influenzato i risultati a causa della quantità limitata di dati a disposizione. Per quanto riguarda il secondo e terzo periodo, in entrambi i casi si può rifiutare l'ipotesi nulla e accettare che esiste un trend negativo. Per quanto riguarda il test Advanced Dickey-Fuller, i risultati ottenuti sono riportati in Tabella 3.7:

Questi valori confermano la non stazionarietà della serie. In particolare, il valore positivo ottenuto per l'ADF Statistic e il valore molto alto di p -value

ADF Statistic	2.364866
p-value	0.998993
1% Crit. Value	-3.889
5% Crit. Value	-3.054
10% Crit. Value	-2.667

Table 3.7: Risultati del test ADF per i tiri da due punti tentati.

suggeriscono che la serie ha una radice unitaria, quindi non si può rifiutare l'ipotesi nulla del test.

Per avere la conferma della non stazionarietà, è stato applicato anche il test KPSS, ottenendo i risultati in Tabella 3.8:

KPSS Statistic	0.662110
p-value	0.016990
10% Crit. Value	0.347
5% Crit. Value	0.463
2.5% Crit. Value	0.574
1% Crit. Value	0.739

Table 3.8: Risultati del test KPSS per i tiri da due punti tentati.

Osservando i risultati ottenuti, si può facilmente notare che, ad esclusione del valore critico con livello di significatività dell'1%, il valore della KPSS Statistic supera tutti i valori critici a tutti i livelli di significatività. Combinando questi dati e il valore del p -value, si può rifiutare l'ipotesi nulla e accettare l'ipotesi alternativa, ovvero che la serie ha una radice unitaria e quindi è non stazionaria attorno a una media costante.

L'analisi della serie tramite l'algoritmo BEAST ha individuato la presenza di due changepoint nella serie temporale, e la tendenza in diminuzione è stata confermata attraverso il test di Mann-Kendall. Ulteriori conferme sulla presenza di trend all'interno della serie sono arrivate attraverso i test ADF e KPSS, che hanno evidenziato una radice unitaria della serie e quindi la non

stazionarietà di quest'ultima.

3.3 Correlazione tra tiri tentati e media punti a partita

Dopo aver rilevato un significativo cambiamento nei trend dei tentativi di tiro nelle squadre NBA, emerge chiaramente una marcata preferenza per i tiri da tre punti nell'era moderna, a scapito dei tiri da due punti. Le squadre attuali mostrano una tendenza a privilegiare strategie di gioco più efficienti, ottimizzando l'utilizzo dei tiri da tre punti. Questo stile di gioco ha portato ad un notevole aumento del ritmo delle partite NBA, caratterizzato da un incremento dei tiri nei primi secondi delle azioni e da punteggi sempre più elevati, simili a quelli dell'All-Star Game.

In questa sezione, si intende verificare se l'incremento dei tiri da tre punti e la diminuzione dei tiri da due punti siano correlati con l'aumento dei punti per partita nelle stagioni NBA. Questo analisi mira a dimostrare, da un punto di vista statistico, l'efficacia dell'ampio utilizzo dei tiri da tre punti rispetto all'approccio predominante del tiro da due punti nel basket degli anni '90.

I dati sono stati suddivisi in due differenti dataframes, consentendo un confronto diretto per valutare l'eventuale correlazione tra questi aspetti. La stagione di riferimento per questa suddivisione è il 2009/10, scelta per garantire una divisione equilibrata dei dati tra i due gruppi.

I due dataframes sono stati esaminati utilizzando sia la funzione di Cross-Correlazione che l'analisi di Granger Causality. La Cross-Correlazione è stata impiegata per valutare le relazioni temporali tra i tiri da tre punti, i tiri da due punti e i punti totali per partita. D'altra parte, l'analisi di Granger Causality è stata utilizzata per determinare se i cambiamenti nei tiri da tre punti e da due punti possano causare variazioni significative nei punti per partita.

3.3.1 Pearson Correlation

Per effettuare questa analisi sono stati calcolati i valori della correlazione di Pearson e le Cross-Correlation come descritte precedentemente. I risultati mostrati nella Tabella 3.9 indicano i risultati ottenuti per i dati riguardanti i tiri da tre punti tentati:

3PTA - PPG	Pearson	Lag0	Lag1	Lag2
Pre 2009/10	0.7488	0.74877	0.581	0.6345
Post 2009/10	0.9345	0.93451	0.923	0.7431

Table 3.9: Risultato della Cross-Correlation tra i tiri da tre punti tentati e i punti segnati per partita.

La tabella evidenzia che, in entrambi i periodi esaminati, esiste una forte correlazione positiva tra i tentativi di tiro da tre punti e i punti segnati per partita dalle squadre NBA. In particolare, la correlazione aumenta nel secondo periodo, dove è stato individuato il changepoint da parte dell'algoritmo BEAST, indicando un impatto ancora maggiore dei tiri da tre punti sui punti segnati. Infine si nota un leggero calo della cross-correlazione nei ritardi temporali (lag 1 e lag 2), ma rimane comunque elevata, suggerendo una persistenza nell'effetto dei tiri da tre punti sull'efficacia nel segnare punti. Per quanto riguarda i tiri da due punti in Tabella 3.10 sono riportati i risultati ottenuti:

2PTA - PPG	Pearson	Lag0	Lag1	Lag2
Pre 2009/10	-0.2534	-0.25341	-0.2609	-0.4147
Post 2009/10	-0.8619	-0.8619	-0.83	-0.7646

Table 3.10: Risultato della Cross-Correlation tra i tiri da due punti tentati e i punti segnati per partita.

In questo caso c'è una correlazione negativa in entrambi i periodi. Nel periodo precedente alla stagione 2009/10, la correlazione è negativa ma debole,

suggerendo che nel in questo periodo c'è un tendenza leggermente negativa tra i tentativi di tiro da due punti e i punti totali segnati per partita. Questo fattore potrebbe indicare che aumentando i tentativi di tiro da due punti potrebbe non portare necessariamente ad un aumento proporzionale nei punti segnati. Al contrario, nelle stagioni successive alla stagione 2009/10, la correlazione è negativa e molto forte, indicando che in questo periodo aumentando i tiri da due punti tentati è associata una diminuzione significativa nel totale dei punti segnati per una partita. Anche i valori ottenuti nei ritardi temporali indicano una persistenza dell'effetto della correlazione negativa, soprattutto nel secondo periodo.

3.3.2 Spearman Correlation

Nell'analisi della relazione tra i tentativi di tiro e i punti segnati per partita nelle squadre NBA, è essenziale esplorare non solo la correlazione diretta, ma anche come questa relazione si manifesta nel tempo. A tale scopo, è stato utilizzato il coefficiente di correlazione di Spearman, il quale valuta la correlazione basandosi sull'ordine dei dati anziché sulla linearità, offrendo così un'analisi robusta anche in presenza di dati non normalmente distribuiti o di relazioni non lineari. I risultati ottenuti con Spearman hanno permesso di esaminare come i cambiamenti nei tiri tentati influenzano i punti segnati nel corso del tempo, evidenziando la persistenza e la forza di questa relazione anche con ritardi temporali.

Nella Tabella 3.11 vengono mostrati i risultati ottenuti per la correlazione di Spearman tra i tiri da tre punti tentati e i punti segnati per partita prima della stagione 2009/10:

La correlazione di Spearman mostra un valore significativo di 0.6813, con un p -value di 0.01034, indicando una correlazione positiva e moderatamente forte tra i due parametri nello stesso periodo. Inoltre, sono stati esaminati anche i ritardi temporali nella correlazione. Sia la correlazione con lag 1 che con lag 2 mostrano valori simili di 0.4727, con p -value di 0.14199 per entrambi i lag. Questi risultati suggeriscono che, mentre la correlazione tra

		<i>p</i> -value
Spearman Correlation	0.6813	0.01034
Lag1 Correlation	0.4727	0.14199
Lag2 Correlation	0.4727	0.14199

Table 3.11: Risultato della correlazione di Spearman tra i tiri da tre punti tentati e i punti segnati per partita prima del 2009/10.

i tiri da tre punti e i punti segnati per partita è meno evidente nei periodi temporali precedenti, persiste comunque una tendenza positiva, seppur meno significativa statisticamente.

Dopo la stagione 2009/10, i risultati riportati in Tabella 3.12 indicano una forte e significativa correlazione positiva tra i tentativi di tiro da tre punti e i punti segnati per partita. I valori ottenuti per Lag1 e Lag2 suggeriscono che la preparazione e la strategia di gioco che favoriscono i tiri da tre punti non solo portano a un aumento dei punti segnati immediatamente, ma potrebbero anche contribuire a un miglioramento delle prestazioni offensive a lungo termine.

		<i>p</i> -value
Spearman Correlation	0.8901	4.56543e-05
Lag1 Correlation	0.9091	0.00010
Lag2 Correlation	0.9091	0.00010

Table 3.12: Risultato della correlazione di Spearman tra i tiri da tre punti tentati e i punti segnati per partita dopo 2009/10.

Per quanto riguarda i tiri da due punti, i risultati per il periodo precedente alla stagione 2009/10 sono mostrati nella Tabella 3.14:

Questi risultati indicano che, prima della stagione 2009/10, non c'era una relazione significativa tra i tentativi di tiro da due punti e i punti segnati per partita. La debole correlazione negativa suggerisce che un aumento dei tentativi di tiro da due punti potrebbe essere associato a una lieve dimin-

		<i>p</i> -value
Spearman Correlation	-0.1978	0.51713
Lag1 Correlation	-0.1182	0.72928
Lag2 Correlation	-0.1182	0.72928

Table 3.13: Risultato della correlazione di Spearman tra i tiri da due punti tentati e i punti segnati per partita prima del 2009/10.

uzione nei punti segnati, ma questa relazione non è supportata da evidenze statistiche solide. Questo potrebbe riflettere l'efficacia complessiva dei tiri da due punti durante questo periodo, dove altre variabili potrebbero aver avuto un'influenza maggiore sul punteggio complessivo delle squadre.

Al contrario, i risultati mostrati in Tabella 3.14 indicano che, dopo la stagione 2009/10, c'è una forte correlazione negativa tra i tentativi di tiro da due punti e i punti segnati per partita. In altre parole, un aumento dei tentativi di tiro da due punti è associato a una significativa diminuzione nei punti segnati. La significatività statistica di queste correlazioni, anche con ritardi temporali, suggerisce che i tentativi di tiro da due punti non solo non sono efficaci immediatamente ma continuano ad avere un effetto negativo sulle prestazioni delle squadre nelle partite successive. Questo rafforza l'idea che le squadre NBA, nel periodo post-2009/10, abbiano ridotto l'uso dei tiri da due punti a favore di strategie di tiro più efficienti, come i tiri da tre punti.

		<i>p</i> -value
Spearman Correlation	-0.8901	4.56543e-05
Lag1 Correlation	-0.9	0.00015
Lag2 Correlation	-0.9	0.00015

Table 3.14: Risultato della correlazione di Spearman tra i tiri da due punti tentati e i punti segnati per partita dopo il 2009/10.

3.3.3 Granger causality

Per determinare se esiste una relazione di causalità Granger tra i tentativi di tiro, sia da due punti che da tre punti, e i punti segnati per partita, sono stati condotti dei test specifici. È importante sottolineare che questi test sono stati eseguiti considerando l'intero dataset senza suddividerlo in periodi temporali distinti, a causa della limitata disponibilità di dati stagionali nel dataset utilizzato.

I risultati del test di Granger Causality per i tentativi di tiro da tre punti rispetto ai punti segnati per partita sono riportati nella Tabella 3.15:

Lag	Test Statistic	<i>p</i> -value
1.0	22.806481	0.000011
2.0	18.842860	0.000295
3.0	27.509099	0.000016

Table 3.15: Risultati del test di Granger Causality tra i tiri da tre punti tentati e i punti segnati per partita.

Questi risultati indicano che i tentativi di tiro da tre punti hanno una relazione causale significativa sui punti segnati per partita delle squadre NBA. Ciò suggerisce che variazioni nell'utilizzo dei tiri da tre punti possono influenzare in modo rilevante il totale dei punti segnati durante le partite.

Analogamente, i risultati del test di Granger Causality per i tentativi di tiro da due punti rispetto ai punti segnati per partita sono presentati nella Tabella 3.16:

Anche in questo caso, i risultati mostrano una relazione causale significativa tra i tentativi di tiro da due punti e i punti segnati per partita. Questo implica che le variazioni nell'uso dei tiri da due punti possono influenzare notevolmente il risultato complessivo delle partite.

In conclusione, i risultati dei test di Granger Causality suggeriscono che entrambi i tipi di tiri, sia da tre punti che da due punti, hanno un effetto causale diretto e significativo sui punti segnati per partita dalle squadre NBA.

Lag	Test Statistic	p -value
1.0	22.655839	1.203226e-05
2.0	19.739001	1.922491e-04
3.0	35.326298	3.980858e-07

Table 3.16: Risultati del test di Granger Causality tra i tiri da due punti tentati e i punti segnati per partita.

3.4 Predizione vittorie

Nel seguente paragrafo, esamineremo i risultati ottenuti dall'analisi della regressione lineare utilizzata per studiare come i fattori considerati nei precedenti paragrafi influenzino la percentuale di vittorie delle squadre NBA. Le performance del modello sono state valutate utilizzando due metriche principali:

- **Mean Squared Error(MSE):** Questa metrica quantifica l'errore quadratico medio tra le predizioni del modello e i valori effettivi della percentuale di vittorie.
- **R2 Score:** Il coefficiente di determinazione R^2 misura la proporzione di varianza nella variabile dipendente (percentuale di vittorie) spiegata dalle variabili indipendenti considerate nel modello.

I valori ottenuti da queste metriche sono riportati in Tabella 3.17:

MSE	0.00133
R2 Score	0.94626

Table 3.17: Risultati MSE e R2 Score per la regressione lineare.

Il modello ha mostrato una buona precisione riguardo all'MSE, indicando la sua capacità di predire in modo accurato la percentuale di vittorie delle squadre NBA. Un valore elevato per il coefficiente R^2 (0.94626) conferma che la variazione nella percentuale di vittorie può essere spiegata efficacemente

dalle variabili indipendenti incluse nella regressione. Inoltre nella Tabella 3.18 sono mostrati i coefficienti stimati dalla regressione lineare:

3PA	0.000189
2PA	-0.000001
ORtg	0.027481
DRtg	-0.029979
Pace	-0.004460
PPG	0.002898

Table 3.18: Coefficienti della regressione lineare

I valori dei coefficienti indicano il peso relativo di ciascuna variabile indipendente nel contribuire alla percentuale di vittorie di una squadra NBA. Dai coefficienti, emerge che i tentativi di tiro da tre punti, l'efficienza offensiva e difensiva, e i punti segnati per partita sono variabili significative nella predizione delle performance di una squadra NBA in termini di percentuale di vittorie.

Questi risultati supportano l'importanza di strategie che potenziano l'efficacia nei tiri da tre punti, migliorano i rating offensivo e difensivo, e aumentano i punti segnati per partita per ottenere migliori risultati nelle competizioni della NBA.

Chapter 4

Conclusioni

Lo studio presentato si propone di analizzare la distribuzione dei tiri tentati da due e tre punti dei giocatori NBA, osservando come le tattiche di gioco siano cambiate nel tempo, con un aumento del ritmo e una maggiore enfasi sul tiro da tre punti. Sono state condotte analisi di correlazione e causalità basate sul numero di tiri tentati dai giocatori NBA. I tiri effettuati durante una stagione, sia da due punti che da tre punti, sono stati utilizzati per comprendere una possibile relazione con i punti segnati per partita.

L'analisi ha evidenziato cambiamenti significativi nelle strategie di tiro delle squadre NBA nel tempo, con un aumento dei tiri da tre punti tentati, particolarmente accentuato dopo la stagione 2010/11. Questi cambiamenti sono stati accompagnati da un aumento della media dei punti segnati per partita. D'altro canto, è emerso un trend negativo per i tiri da due punti, indicando una possibile diminuzione della loro importanza nelle strategie di gioco moderne.

I test di stazionarietà effettuati hanno evidenziato la presenza di radici unitarie nelle serie temporali analizzate, indicando che le serie risultano essere non stazionarie. Applicando il test di causalità di Granger, si è osservata una relazione di precedenza temporale tra i tiri tentati e la media dei punti segnati, suggerendo che i trend nei tiri possono essere indicatori anticipatori di una maggiore o minore produzione offensiva.

Infine, il modello di regressione lineare ha fornito ulteriori informazioni sull'influenza dei diversi fattori sulla percentuale di vittorie delle squadre NBA. In particolare, i tentativi di tiro da tre punti, insieme all'efficienza offensiva e difensiva, sono emersi come variabili significative nel determinare le performance delle squadre.

Sulla base dei risultati ottenuti, è possibile rispondere alle domande di ricerca come segue:

- Il numero di punti segnati per partita è fortemente correlato in modo positivo con i tiri tentati da tre punti e correlato negativamente con i tiri da due punti tentati.
- Fattori come punti segnati per partita, tiri tentati da tre punti, efficienza difensiva e offensiva risultano essere buoni predittori delle performance di una squadra NBA.
- A livello statistico, il tiro da tre punti risulta essere più efficiente del tiro da due punti.

Il trend positivo dei tiri tentati da tre punti iniziato a partire dal 2000 è una conseguenza dei cambiamenti regolamentari effettuati dalla NBA, come l'ammissione della difesa a zona. Successivamente, nella stagione 2010/11, questo fenomeno è esploso con l'arrivo di Stephen Curry e Klay Thompson ai Golden State Warriors. Il duo, soprannominato "Splash Brothers", ha basato il proprio gioco sul tiro da tre punti, influenzando molte squadre NBA che hanno cercato di replicare questa tattica. Allo stesso tempo, il trend negativo dei tiri tentati da due punti è iniziato negli anni 2000, a causa delle nuove regole introdotte nella NBA. Questo trend è stato ulteriormente supportato da Daryl Morey, che con l'introduzione della "Moreyball" ha creato un sistema di gioco altamente efficiente basato esclusivamente su tiri da tre punti o da due punti nei pressi del ferro. Gli Houston Rockets hanno estremizzato questa strategia, avvicinandosi a una finale NBA seguendo questa tattica.

4.1 Limiti dello studio

Nonostante i risultati ottenuti siano significativi, questo studio presenta alcune limitazioni che devono essere considerate. Innanzitutto, l'analisi è basata su serie temporali aggregate a livello di stagione, il che potrebbe mascherare variazioni più sottili che avvengono durante singole partite o in periodi più brevi della stagione. Questo livello di aggregazione limita la capacità di catturare dinamiche di breve termine che potrebbero essere rilevanti.

In secondo luogo, l'utilizzo di soli due test di stazionarietà (ADF e KPSS) potrebbe non essere sufficiente per garantire una robusta verifica della stazionarietà delle serie temporali. Altri test, come il test di Phillips-Perron, potrebbero fornire ulteriori conferme o contraddizioni rispetto ai risultati ottenuti. Un'altra limitazione riguarda l'applicazione del test di causalità di Granger. Questo test, pur essendo potente, non implica una vera causalità ma solo una relazione di precedenza temporale. Pertanto, i risultati del test di Granger devono essere interpretati con cautela e non possono essere considerati definitivi per stabilire relazioni causali tra variabili.

Inoltre, l'analisi si concentra esclusivamente su un insieme limitato di variabili indipendenti (tiri da due e da tre punti, ritmo di gioco, rating offensivo e difensivo, e punti segnati per partita). Potrebbero esistere altre variabili significative, come la qualità del roster, gli infortuni dei giocatori chiave, o le strategie degli allenatori, che non sono state considerate ma che potrebbero influenzare significativamente la percentuale di vittorie di una squadra NBA. Infine, l'algoritmo BEAST utilizzato per l'individuazione dei changepoint è avanzato ma complesso, e le sue assunzioni e impostazioni possono influenzare i risultati. Errori nella selezione dei parametri del modello o nella interpretazione dei risultati potrebbero portare a conclusioni fuorvianti. È quindi importante considerare questi risultati nel contesto di altre tecniche e verificare la robustezza dei changepoint individuati con metodi alternativi.

Queste limitazioni evidenziano la necessità di ulteriori ricerche che possano approfondire e migliorare la comprensione delle dinamiche analizzate in questo

studio, integrando dati più dettagliati e metodologie complementari.

4.2 Conclusioni

In conclusione, il luogo comune secondo cui la NBA è caratterizzata sempre più da giocatori capaci di segnare da qualunque posizione e distanza, e che da ormai molti anni è orfana dei "centri di una volta", è in parte confermato. Infatti, si è dimostrato che, a livello statistico, il tiro da tre punti è più influente nel determinare il punteggio delle partite NBA.

Gli effetti del tiro da tre punti costituiscono una correlazione positiva con il numero di punti segnati per partita. D'altro canto, i tiri tentati da due punti mostrano una correlazione negativa, sebbene in maniera meno marcata. Questo studio conferma l'importanza crescente dei tiri da tre punti nelle strategie di gioco moderne e il loro impatto significativo sulle prestazioni delle squadre NBA.

In sintesi, l'analisi delle serie temporali e dei modelli statistici applicati ha evidenziato cambiamenti significativi nelle strategie di tiro delle squadre NBA, con un aumento rilevante dei tiri da tre punti tentati negli ultimi due decenni. La correlazione e causalità tra tiri tentati e punti segnati per partita indicano che le squadre moderne beneficiano notevolmente di una maggiore enfasi sul tiro da tre punti. Infine, i modelli di regressione lineare hanno mostrato che variabili come i tiri tentati da tre punti e l'efficienza offensiva e difensiva sono determinanti cruciali per la percentuale di vittorie delle squadre NBA.

4.3 Lavori Futuri

I risultati di questo studio aprono la strada a diverse possibilità di approfondimento e miglioramento per future ricerche. Di seguito sono delineate alcune direzioni potenziali per ulteriori indagini.

- Un'analisi più dettagliata potrebbe esaminare non solo la quantità di tiri tentati, ma anche la qualità dei tiri e le situazioni di gioco in cui

vengono effettuati. Ad esempio, si potrebbero studiare le differenze tra i tiri tentati in situazioni di pick-and-roll, isolamento o transizione, e come queste influiscano sull'efficienza offensiva complessiva.

- Mentre questo studio si è concentrato principalmente sugli aspetti offensivi del gioco, future ricerche potrebbero integrare un'analisi più robusta dei fattori difensivi. Esaminare come le strategie difensive influiscano sulla scelta dei tiri e sulla loro efficacia potrebbe fornire una visione più completa delle dinamiche di gioco.
- L'implementazione di modelli di machine learning potrebbe migliorare le previsioni sulle performance delle squadre e l'efficacia delle strategie di tiro. Algoritmi di apprendimento automatico possono identificare pattern complessi e offrire raccomandazioni strategiche basate su grandi volumi di dati.
- Dato il cambiamento nel ruolo dei centri evidenziato in questo studio, future ricerche potrebbero focalizzarsi esclusivamente su questa posizione. Analizzare come i centri moderni contribuiscono alle strategie di tiro e alle dinamiche di squadra potrebbe fornire insights preziosi sulle evoluzioni tattiche in corso.

Bibliography

1. Issa, M. *The 7-Seconds-or-Less Suns: An Offensive Dynasty for the Ages* <https://theanalyst.com/eu/2022/07/the-seven-seconds-or-less-phoenix-suns/>.
2. Buckley, Z. *Visualizing Splash Brothers' Climb Up NBA's All-Time 3-Point List* <https://bleacherreport.com/articles/2391384>.
3. Lawyer, M. The Evolution of the Center. *On The Button*. <https://onthebutton.net/2024/03/15/evolution-of-the-center-nba-history/> (2024).
4. Sanders, M. G. *et al.* Evolution of the Center: How the Introduction of the Three-Point Shot Devalued the Traditional Big Man and Altered the Offensive Profile of NBA Centers. *Journal of Sports Analytics*. <https://dsi.dartmouth.edu/journal> (2022).
5. Deegan, T. Data Analysis on The NBA's Three-Point Revolution. *NYC Data Science Academy Blog*. <https://nycdatascience.com/blog/student-works/data-analysis-on-the-nbas-three-point-revolution/> (2020).
6. Williams, L. & White, S. The Rise of Stephen Curry and the Golden State Warriors. *International Journal of Sports History* **8**, 456–470 (2017).
7. Ehrlich, J. *et al.* Estimating NBA Team Shot Selection Efficiency from Aggregations of True, Continuous Shot Charts: A Generalized Additive

- Model Approach. *SSRN Preprint*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4697111 (2024).
8. NBA. *NBA Draft Lottery: How it works* <https://www.nba.com/news/nba-draft-lottery-explainer>. Accessed: 2024-06-02. 2023.
 9. Foundation, P. S. *Requests: HTTP for Humans* Accessed: 2024-06-02. 2024. <https://pypi.org/project/requests/>.
 10. Richardson, L. *Beautiful Soup 4* Accessed: 2024-06-06. 2024. <https://pypi.org/project/beautifulsoup4/>.
 11. Brockwell, P. J. & Davis, R. A. *Introduction to time series and forecasting* (Springer, 2016).
 12. Zhao, K. *et al.* Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment* **232**, 111181. ISSN: 0034-4257. <https://www.sciencedirect.com/science/article/pii/S0034425719301853> (2019).
 13. Hussain, M. M. & Mahmud, I. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* **4**, 1556. <https://doi.org/10.21105/joss.01556> (2019).
 14. Hamed, K. H. & Ramachandra Rao, A. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* **204**, 182–196. ISSN: 0022-1694. <https://www.sciencedirect.com/science/article/pii/S002216949700125X> (1998).
 15. Dickey, D. & Fuller, W. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *JASA. Journal of the American Statistical Association* **74** (June 1979).
 16. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438 (1969).

17. Lütkepohl, H. *New introduction to multiple time series analysis* (Springer Science & Business Media, 2005).
18. Derrick, T. in, 189–205 (Jan. 2004). ISBN: 9780736044677.
19. Shojaie, A. & Fox, E. B. *Granger Causality: A Review and Recent Advances* 2021. arXiv: 2105.02675.
20. ESPN. *NBA approves new rules* Accessed: 2024-06-06. 2001. <http://static.espn.go.com/nba/news/2001/0412/1172176.html%5C#:~:text=The%5C%20changes%5C%20are%5C%3A,is%5C%20not%5C%20closely%5C%20guarding%5C%20anyone..>
21. Thoele, C. A. *Moneyball: The Effects of Three-Point Shooting and Small-Ball Lineups in the NBA* Accessed: 2024-06-09. 2022. https://fisherpub.sjf.edu/cgi/viewcontent.cgi?article=1106&context=sport_undergrad#:~:text=Moreyball%20is%20the%20concept%20of,the%20coaches%20of%20the%20NBA..