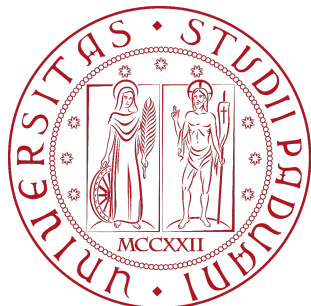


Università degli studi di Padova
Dipartimento di Scienze Statistiche

Corso di Laurea Magistrale in
Scienze Statistiche



VALUTAZIONE STATISTICA DELLE PERFORMANCE
DEI GIOCATORI DELLA NATIONAL BASKETBALL
ASSOCIATION

Relatore: Prof. Nicola Sartori

Dipartimento di Scienze Statistiche

Laureando: Davide Artuso

Matricola n°:1210594

Anno Accademico 2019/2020

Indice

Introduzione	5
1 I dati	9
1.1 Fonte dei dati	9
1.2 Grandi masse di dati	10
1.3 Creazione del <i>dataset</i> finale	12
1.3.1 Matrice delle variabili esplicative	13
1.3.2 Variabile risposta	16
1.4 Principali indici statistici utilizzati	17
2 Modello logistico per la valutazione dei giocatori NBA	23
2.1 Modello di regressione logistica	24
2.2 Problemi del modello e correzione di Firth	31
2.3 Costruzione del <i>ranking</i>	33
2.3.1 Primi metodi e problemi computazionali	34
2.3.2 La statistica EWMA per aggregare i risultati	35
2.3.3 Calcolo del <i>ranking</i>	40
2.4 Risultati	41
2.4.1 Valutazione globale dei giocatori	41
2.4.2 Valutazione dei giocatori stratificata per ruolo	44
2.4.3 Valutazione delle performance di squadra	48
2.5 Valutazioni finali	50
3 Stima bayesiana del modello logistico per la valutazione dei giocatori NBA	53
3.1 L'inferenza <i>bayesiana</i>	54

3.1.1	<i>Markov Chain Monte Carlo</i>	57
3.2	<i>Hamiltonian Monte Carlo</i>	57
3.2.1	Limiti dei metodi precedenti	58
3.2.2	Il metodo generale	58
3.3	Inferenza <i>bayesiana</i> nel modello logistico	61
3.3.1	Scelta delle distribuzioni a priori e calcolo della distri- buzione a posteriori	61
3.3.2	Stima del modello tramite <i>HMC</i>	62
3.3.3	STAN	63
3.3.4	L'importanza delle distribuzioni a priori	64
3.4	Costruzione del <i>ranking</i>	65
3.4.1	Primi metodi utilizzati	65
3.4.2	Un vantaggio dei metodi <i>bayesiani</i>	67
3.4.3	Inserimento dell'effetto fisso	69
3.5	Valutazione delle performance di alcuni giocatori	72
3.5.1	Valutazione globale dei giocatori	76
3.5.2	Valutazione dei giocatori stratificata per ruolo	77
3.5.3	Valutazione delle performance di squadra	79
3.6	Un modello gerarchico con effetti annidati	81
3.7	Valutazioni finali	86
	Conclusioni	89
	Bibliografia	93
	A Codice R utilizzato	97
	B Tabelle dettagliate	115

Introduzione

La statistica è una disciplina che interessa svariati settori, dall'economia alla medicina, dalla politica alle nuove tecnologie. Questa scienza sta prendendo sempre più piede e negli ultimi anni, con l'aumento della disponibilità di grandi masse di dati e delle capacità di calcolo dei computer, gli ambiti applicativi che la coinvolgono sono aumentati a dismisura. Un ulteriore settore in cui questa disciplina viene applicata, e che recentemente sta acquisendo grande popolarità, è quello sportivo, ed è proprio questo l'ambito nel quale si colloca la presente tesi. In moltissimi sport infatti, come ad esempio calcio, basket e pallavolo, vengono raccolti sempre più dati, al fine di analizzarli per fornire delle indicazioni riguardo lo sport considerato. Le analisi che vengono svolte possono avere molteplici obiettivi, come ad esempio il calcolo delle quote nel mondo delle scommesse o la valutazione della qualità delle squadre o dei singoli giocatori, per dare delle indicazioni sull'appetibilità degli stessi a livello di mercato.

Uno degli sport che predilige maggiormente l'utilizzo della statistica è il basket. Nella pallacanestro infatti possono essere raccolte una grandissima quantità di informazioni dalle quali si possono ricavare moltissimi indici. Questi vengono poi analizzati per valutare la bravura di un giocatore, o ancora per stabilire come determinate tipologie di gioco o di strategie riescano ad influenzare l'esito della partita. La quasi totalità dei metodi utilizzati per fornire queste valutazioni è tuttavia di tipo descrittivo. Tra le statistiche più rilevanti si trovano: la percentuale di tiri da due/tre punti realizzati sul totale di tiri da due/tre punti effettuati, la percentuale di tiri liberi realizzati, il numero di palle perse e recuperate, alcuni indici riguardanti le stoppate, dove per stoppata si intende il gesto tecnico con il quale il giocatore in difesa

devia un tiro a canestro del giocatore in attacco, e il numero di rimbalzi, ovvero il numero di palloni recuperati in seguito ad un tiro sbagliato.

Oltre agli indici descrittivi tuttavia, con la mole di dati a disposizione, si può pensare anche di utilizzare dei metodi per modellare i dati in modo da fornire delle valutazioni probabilistiche, considerando contemporaneamente più caratteristiche del giocatore o della giocata effettuata, in modo tale da comprendere quali siano le giocate più prolifiche per la realizzazione di un canestro, o quali siano le tecniche difensive che più probabilmente consentono di evitare il tiro da parte della squadra avversaria, permettendo anche di effettuare delle previsioni sulle partite da giocare.

L'obiettivo di questa tesi è quello di utilizzare varie tecniche statistiche per valutare le performance offensive dei giocatori di NBA (National Basketball Association). Lo scopo è costruire un *ranking* dei giocatori stessi, ovvero una classifica dei giocatori che permetterà di valutare come nel complesso questi si sono comportati durante un'intera stagione, permettendo quindi delle valutazioni globali ma anche consentendo di confrontare giocatori che appartengono allo stesso ruolo. Queste valutazioni potrebbero risultare particolarmente utili sia per le squadre nelle quali militano i giocatori nell'arco della stagione, sia in ottica di mercato, per comprendere al meglio quali sono i giocatori più in forma e soprattutto per quali caratteristiche hanno brillato, consentendo di scegliere su quali giocatori puntare per la stagione successiva.

La tesi si organizza come segue: nel Capitolo 1 vengono presentati i dati utilizzati nell'analisi, riguardanti la stagione NBA 2018/2019. In particolare, dopo aver presentato brevemente le variabili coinvolte nello studio, viene trattata la creazione della matrice del disegno e la costruzione della variabile che verrà utilizzata come risposta nei modelli di regressione considerati nei capitoli successivi. Sempre in questo capitolo vengono poi presentati degli indici descrittivi molto usati nel basket che permetteranno di fornire delle prime valutazioni sulle performance offensive dei giocatori.

Nel Capitolo 2 viene utilizzato un modello di regressione logistica, il quale tuttavia presenta alcuni problemi di stima dal punto di vista frequentista dovuti alla quasi perfetta separazione delle variabili. Viene quindi introdotto un differente metodo per la stima dei parametri che utilizza una correzione per la distorsione denominata correzione di Firth (Firth, 1993). Dopo aver

adattato il modello viene quindi calcolato il *ranking* dei giocatori, che permette appunto di stabilire quali sono i giocatori NBA più prolifici a livello offensivo nell'arco della stagione considerata. Infine vengono quindi presentati i risultati ottenuti andando a valutare i singoli giocatori e alcune azioni da loro compiute.

Nel Capitolo 3 viene poi considerato un approccio *bayesiano* alla stima del modello logistico. Dopo una breve introduzione alla statistica *bayesiana* vengono presentati i principali metodi che vengono utilizzati per stimare i parametri e le loro distribuzioni. In particolare, viene introdotto l'*Hamiltonian Monte Carlo* (Duane *et al.*, 1987), un metodo molto utilizzato negli ultimi anni per la stima di modelli *bayesiani* più complessi. Infine, come nel capitolo precedente viene calcolato il *ranking* dei giocatori sulla base del modello stimato con questa metodologia e vengono quindi valutate le performance offensive dei giocatori e alcune azioni da loro effettuate.

Tutte le analisi svolte nella tesi sono state fatte usando la versione 4.0.2 di R (R Core Team, 2020). Il codice utilizzato è riportato nell'Appendice A.

Capitolo 1

I dati

I dati per lo studio e la valutazione delle performance offensive dei giocatori di basket necessitano di un buon livello di dettaglio e devono consentire una valutazione completa di ciascun giocatore, tenendo conto di tutte le giocate che questi effettuano nel corso di ogni partita. Di conseguenza, quello di cui si ha bisogno in un'analisi di questo genere è un cosiddetto *dataset play by play*, ovvero un insieme di dati molto ampio che contiene le informazioni di tutte le azioni effettuate dai singoli giocatori durante tutte le partite disputate. Per azione si intende qualsiasi tocco di palla effettuato dal giocatore, che sia un passaggio, un tiro, una stoppata o una ripartenza. Con insiemi di dati di questo genere è possibile andare ad analizzare diverse caratteristiche dei giocatori e fornire delle valutazioni sulle loro performance al termine di ogni partita disputata.

1.1 Fonte dei dati

Con il passare degli anni sono sempre maggiori le possibilità di ottenere insiemi di dati complessi in tempo reale, o immediatamente dopo la fine di una partita, favorendo l'opportunità di effettuare svariate tipologie di analisi. Sono infatti presenti alcune aziende che si occupano di analizzare in dettaglio le partite della maggior parte degli sport, con lo scopo di ottenere le informazioni riguardo tutte le azioni compiute da ciascun giocatore nel corso di ogni partita. I dati utilizzati nella presente tesi sono stati acquistati nel si-

to BigDataBall (BigDataBall, 2007), il quale fornisce grandi quantità di dati riguardanti lo sport. Il sito citato in precedenza, è gestito dal rinomato NBA-stuffer (NBAsuffer, 2007), un sito web di riferimento che fornisce contenuti analitici. Questi siti mettono a disposizione degli utenti insiemi di dati molto ampi per vari sport molto popolari negli Stati Uniti, come il *football* e appunto il basket, e permettono di ottenere *dataset* contenenti informazioni relative alle statistiche di gioco di ciascuna squadra, oppure alle statistiche dei singoli giocatori. Infine, questi siti permettono di acquistare i dati che sono necessari per affrontare il problema oggetto di studio della presente tesi, ovvero dei *dataset play by play*, per ciascuna stagione NBA dal 2004 in poi. I dati procurati dai precedenti siti ed utilizzati nelle successive analisi riguardano la 73^a edizione del campionato NBA, quella disputata negli anni 2018/2019 e conclusasi con la vittoria del titolo da parte dei Toronto Raptors.

1.2 Grandi masse di dati

Con insiemi di dati della tipologia descritta nel paragrafo precedente, si ha a disposizione una quantità di informazione particolarmente elevata. Avendo infatti per ogni partita disputata nel corso dell'anno tutte le informazioni relative alle azioni compiute da ciascun giocatore, si ha che il numero di osservazioni, ovvero il numero totale di azioni effettuate nell'arco di un'intera stagione, risulterà particolarmente elevato.

Il numero di osservazioni tuttavia non è l'unico aspetto di cui tenere conto, i dati forniti dal sito BigDataBall infatti, forniscono un livello di dettaglio notevole anche per quanto riguarda la tipologia di giocata effettuata, mettendo quindi a disposizione un numero di variabili abbastanza elevato per la descrizione dell'azione. Quando si ha a che fare con insiemi di dati di questa tipologia non si può dunque fare a meno di parlare di *Big Data*, un termine sempre più popolare nell'ambito della statistica e del *Data Science*. Con *Big Data* ci si riferisce ad un insieme di dati prodotti in grande quantità, con una notevole rapidità e nei formati più diversi, la cui elaborazione richiede tecnologie e risorse che vanno ben al di là dei convenzionali sistemi di gestione dei dati (Rezzani, 2013).

Per costruire un sistema che permetta la valutazione delle performance dei giocatori attraverso la costruzione di un *ranking* dei giocatori stessi, viene quindi utilizzato un insieme di dati molto ampio riguardanti la stagione menzionata in precedenza. Questi hanno un peso in memoria di 194 MB e mettono a disposizione un numero pari a circa 700 mila azioni, con un'elevata quantità di variabili che verranno descritte nel seguito, indicanti le varie tipologie di azione compiute dai giocatori e altre caratteristiche degli stessi. Di conseguenza, per ciascuna azione di gioco si ha a disposizione:

- **game_id**: codice identificativo del match;
- **data_set**: variabile che indica se si tratta di una partita della "Regular Season" o dei "Playoff";
- **date**: data nella quale si è giocata la partita;
- **a1, a2, a3, a4, a5**: variabili che riportano i cinque giocatori in campo per la squadra fuori casa;
- **h1, h2, h3, h4, h5**: variabili che riportano i cinque giocatori in campo per la squadra in casa;
- **period**: quarto in cui si sta svolgendo l'azione considerata;
- **away_score, home_score**: punteggio rispettivamente delle squadre fuori casa e in casa al momento dell'azione considerata;
- **remaining_time, elapsed, play_length**: tre variabili per descrivere l'azione nel corso periodo considerato: rispettivamente corrispondono al tempo rimasto da giocare in quel quarto, al tempo trascorso dall'inizio del quarto e alla lunghezza dell'azione considerata;
- **team**: squadra responsabile dell'azione descritta;
- **event_type**: tipologia di evento o di giocata che si verifica nel periodo di tempo indicato;
- **assist, block**: due variabili che indicano rispettivamente quali giocatori hanno effettuato l'assist per un'eventuale canestro o una stoppata difensiva;

- **player**: giocatore responsabile dell'azione;
- **points**: numero di punti derivanti dalla giocata (se questa corrisponde ad un tiro a canestro);
- **result**: risultato del tiro a canestro;
- **steal**: variabile indicante il giocatore che ha rubato la palla;
- **type**: variabile che fornisce un livello di dettaglio maggiore per la giocata effettuata, consentendo di estendere i concetti espressi dalla variabile `event_type`;
- **shot_distance**: variabile che indica la distanza dal canestro al momento del tiro;
- **original_x, original_y, converted_x, converted_y**: posizione in cui si trova il giocatore al momento del tiro;
- **description**: descrizione dettagliata dell'azione considerata;

1.3 Creazione del *dataset* finale

In un'analisi statistica che prevede la modellazione dei dati con lo scopo di valutare le performance offensive dei giocatori, la prima cosa da fare sono una serie di operazioni di *preprocessing*, che consentiranno di utilizzare l'insieme di dati per affrontare il problema in questione.

Prima di tutto quindi, come verrà approfondito nel paragrafo seguente, sono state studiate attentamente le singole variabili descritte in precedenza, per comprendere quali potessero essere utili nel modellare le qualità delle giocate offensive dei giocatori, e quindi per capire quali inserire nel *dataset* finale. Dopodiché sono state effettuate una serie di operazioni di pulizia e di correzione degli errori presenti nei dati, principalmente dovuti ad imprecisioni in fase di imputazione.

Infine, è stato affrontato il problema di come valutare se un'azione svolta da un giocatore potesse essere pericolosa dal punto di vista offensivo o meno. Questo aspetto infatti, non è descritto all'interno di una variabile, e

può essere considerato come una variabile latente, proprio perché il suo esito non è direttamente osservato, ma può essere dedotto attraverso l'osservazione delle altre variabili. Sarà proprio quest'ultima quindi a definire l'importanza di un'azione offensiva, fungendo da variabile risposta in modelli di regressione che permetteranno di studiare la pericolosità delle azioni compiute dai giocatori NBA, consentendo poi la stima delle performance dei giocatori.

1.3.1 Matrice delle variabili esplicative

Come descritto all'inizio della sezione, la prima cosa fatta per il raggiungimento dell'obiettivo, è stata creare la matrice delle variabili esplicative, ovvero una matrice $n \times p$ dove a ciascuna riga corrisponde un'azione e a ciascuna colonna una variabile che la descrive.

Per la creazione di questa matrice, sono state inizialmente effettuate delle operazioni di preprocessing, volte ad individuare eventuali errori ed incongruenze nei dati. Di conseguenza, sono state eliminate tutte le righe, e quindi le azioni, non associate a nessun giocatore, e successivamente, sono state eliminate le giocate che si riferivano a giocatori non presenti sul terreno di gioco.

Dopo queste operazioni ed altri controlli sui dati, c'è stata un'altra operazione particolarmente importante da svolgere, ovvero l'inserimento del ruolo dei giocatori. Quest'ultimo non era infatti presente nel *dataset* fornito da BigDataBall, ma rappresenta un aspetto importante per l'indagine, poiché consente di svolgere analisi stratificate per ruolo, portando in seguito alla valutazione dei migliori giocatori in ciascun ruolo considerato. Per inserirlo all'interno del *dataset*, è stato utilizzato un altro insieme di dati contenente tutti i giocatori associati ai rispettivi ruoli, e dopodiché è stata effettuata un'operazione di unione per integrare i due dataset mediante l'utilizzo del pacchetto R `dplyr` (Wickham *et al.*, 2020). A questo punto, per ogni azione effettuata da ciascun giocatore, sarà presente anche il ruolo relativo al giocatore stesso, ovvero uno tra: centro (C), playmaker (PG), guardia tiratrice (SG), ala piccola (SF) o ala grande (PF).

Dopo queste operazioni, l'ultimo passo per creare una matrice delle variabili esplicative utilizzabile per valutare le performance offensive dei giocatori

è quello di decidere quali variabili utilizzare. Essendo che l'obiettivo è quello di valutare le giocate aventi lo scopo di realizzare un canestro, il *dataset* finale dovrà includere tutte quelle variabili che descrivono la tipologia di azione effettuata e quelle che danno una descrizione del giocatore. Di conseguenza, sono state rimosse le variabili che avevano lo scopo di descrivere la partita, come le formazioni schierate dalle varie squadre, gli indici temporali indicanti il periodo di gioco nel quale è stata effettuata l'azione e il punteggio delle squadre considerate nel corso del match. Oltre a queste, sono state rimosse anche le variabili relative alle coordinate dei giocatori, poiché presenti solamente per le azioni che descrivono un tiro a canestro, e la descrizione dell'azione, poiché risulta essere un'informazione ridondante dato che è già descritta dalle altre variabili.

A questo punto quindi, per far sì che ogni riga contenga tutte le informazioni relative all'azione svolta e per facilitare le operazioni di modellazione, è stato deciso di trasformare le variabili relative agli eventi considerati e quelle che si riferiscono ai sotto-eventi in variabili indicatrici, ovvero in più covariate che assumono il valore 1 se quella tipologia di evento o sotto-evento si è verificata nell'azione considerata e valore 0 altrimenti. Nella Tabella 1.1 vengono elencati tutti gli eventi e i sotto-eventi ad essi associati che descrivono la tipologia di azione effettuata dai cestisti.

Lo scopo principale della tesi è comunque valutare le performance offensive dei giocatori, e pertanto non è di particolare interesse considerare le categorizzazioni dettagliate riguardanti i falli, le sostituzioni, le violazioni minori, le espulsioni e l'ordine numerico dei tiri liberi. Di conseguenza, tutti questi aspetti non verranno considerati nell'analisi, ma ci si concentrerà principalmente sui tiri, sulle palle perse, sulle palle recuperate, sui tiri liberi e sugli *assist*, ovvero sulle principali categorie di eventi che sono indice di pericolosità e di incisività dei giocatori in fase offensiva e che possono portare alla realizzazione di più o meno canestri.

Una volta effettuate tutte queste operazioni si arriva dunque ad avere una matrice del disegno pronta per essere utilizzata, la quale contiene tutte le informazioni necessarie riguardanti le azioni svolte nel corso delle partite. Come ultimo aspetto, per evitare di dare delle valutazioni affrettate su giocatori che non hanno avuto (nella stagione in questione) il tempo di esprimersi,

Tabella 1.1: Eventi e sotto-eventi relativi alle giocate effettuate

Evento	Sotto-Evento
jump ball: tipologia di azione per cominciare o ricominciare il gioco	jump ball
miss: tiro a canestro sbagliato	39 tipologie di tiro tra cui: Layup Jump Shot Hook Shot
rebound: recupero del pallone dopo un tiro sbagliato	rebound rebound offensive rebound defensive
turnover: palla persa	25 tipologie di giocate che portano alla perdita del pallone tra cui: bad pass 3 second violation lost ball
shot: tiro a canestro realizzato	40 tipologie di tiro tra cui: Layup Jump Shot Hook Shot
foul: fallo commesso da un giocatore	15 tipologie di fallo tra cui: foul away from play foul flagrant.foul.type1
sub: sostituzione del giocatore	sub
free throw: tiro libero	13 tipologie di tiro libero rappresentanti l'ordine del tiro sul totale dei tiri liberi assegnati
violation: falli minori	6 tipologie di violazione tra cui: defensive goaltending jump ball kicked ball
ejection: espulsione	ejection
assist: passaggio precedente al canestro	assist

sono stati rimossi dal dataset tutti quei giocatori che nel corso di tutta la stagione hanno partecipato a meno di 100 azioni. Dopo tutte queste operazioni preliminari e di risoluzione degli errori la matrice finale contiene un numero di righe (e quindi di azioni) pari a 663956 e un numero di variabili che descrivono la giocata pari a 63.

1.3.2 Variabile risposta

Venendo ora alla variabile risposta, questa dovrà permettere di comprendere quali sono le azioni utili da un punto di vista offensivo e quali invece non lo sono. Tuttavia questa variabile non è direttamente osservabile e potrebbe essere frutto di valutazioni soggettive, portando quindi alla necessità di trovare un metodo oggettivo basato sui dati per la sua definizione.

Essendo che l'obiettivo del basket è quello di effettuare più canestri della squadra avversaria, un buon metodo per comprendere se una giocata eseguita può essere considerata utile a livello offensivo, è quello di valutare se questa ha portato, con lo sviluppo dell'azione ed entro i 24 secondi disponibili, alla realizzazione del canestro. Seguendo questa idea, utilizzata anche per sport come il calcio da Decross (Decroos, 2019) o da Dandolo (Dandolo, 2019), è quindi stata creata una variabile dicotomica y , la quale assume valore 1 se l'azione considerata ha portato la squadra alla realizzazione di un canestro nelle successive k azioni e valore 0 altrimenti.

Dopo il calcolo si avrà dunque una variabile dicotomica per ogni azione, che differisce a seconda del ritardo temporale k utilizzato, la quale segue una distribuzione di Bernoulli con parametro ignoto. Il vettore rappresentante la variabile risulta essere del tipo:

$$\mathbf{y}_k = (y_{k,1}, y_{k,2}, \dots, y_{k,n}) \quad y_{k,i} \sim \text{Bernoulli}(\pi_i);$$

Nella definizione precedente è evidenziata la dipendenza della variabile risposta dal *lag* temporale k selezionato, sottolineando appunto l'importanza di questo valore. Al variare di quest'ultimo, la valutazione riguardo alle singole azioni può cambiare anche di molto, infatti, utilizzando k piccolo, verranno selezionate come importanti solamente le azioni immediatamente precedenti alla realizzazione di un canestro, mentre utilizzando k grande, verranno

Tabella 1.2: Vettore rappresentante la variabile risposta

0	1
416402	214590

selezionate come rilevanti una grande quantità di azioni. Nella scelta, bisogna quindi trovare un compromesso tra questi due aspetti. Considerando che nella pallacanestro le azioni sono molto rapide e devono essere concluse per regolamento entro 24 secondi, non sembra particolarmente sensato lavorare con un valore k grande. Dopo aver valutato i risultati ottenuti per vari valori di k , in particolare considerando valori compresi tra 2 e 7, è quindi stato scelto di utilizzare un valore del *lag* temporale pari a 4, il quale ha portato alla suddivisione delle azioni riportata in Tabella 1.2.

Dopo aver costruito la variabile risposta sfruttando le informazioni contenute nel *dataset* ottenuto dopo le operazioni preliminari, si ottiene quindi un insieme di dati completo, pulito ed utilizzabile per analizzare l'efficacia delle giocate dei cestisti. La Tabella 1.3 riporta le prime righe del *dataset* dopo la fase di *preprocessing*.

1.4 Principali indici statistici utilizzati

In seguito alla costruzione del dataset, vengono qui presentati alcuni indici descrittivi ampiamente utilizzati in NBA e nel mondo della pallacanestro per fornire delle valutazioni statistiche dei giocatori. Nel corso degli anni, vari esperti hanno infatti proposto diversi metodi per valutare i giocatori combinando degli indicatori riassuntivi delle loro performance. In particolare, gli indici descritti nel seguito sono stati calcolati per ciascun giocatore a disposizione:

- **PG** e **MIN**: numero di partite giocate e corrispettivo minutaggio in campo
- **AS**: numero totale di assist effettuati

Tabella 1.3: Prime righe del *dataset* utilizzato nelle analisi dopo le operazioni preliminari

date	game_id	data_set	player	team	foul	free_throw
2018-10-16	0021800001	2018-19 Regular Season	Al Horford	BOS	0	0
2018-10-16	0021800001	2018-19 Regular Season	Robert Covington	PHI	0	0
2018-10-16	0021800001	2018-19 Regular Season	Jayson Tatum	BOS	0	0
jump_ball	rebound	shot	turnover	violation	AlleyOpD	AlleyOpL
1	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	1	0	0	0	0
backcourt	backpass	discourtnueDribble	doubleDribble	DrivingBankHook	DrivingFingerL	DrivingHook
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
DrivingReversel	Dunk	FadeawayJump	FingerRoll	FloatingJump	HookBank	HookShot
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
jumpball.1	JumpBank	JumpShot	Layup	lostBall	offensiveGoaltending	outofBoundsLB
1	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	1	0	0	0	0
palmingT	PullupJump	PutbackDunk	PutbackLayup	Rdefensive	Roffensive	ReverseDunk
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
ReverselLayup	RunningDunk	RunningFingerL	RunningJump	RunningLayup	RunningReversel	StepBackJump
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
stepoutofBounds	traveling	TurnaroundBankHook	TurnaroundFadeaway	TurnaroundHook	TurnaroundJump	turnover.1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
unknown	assist	steal	block	ruolo	res	y
0	0	0	0	C	0	0
0	0	0	0	SF	0	0
0	0	0	0	SF	0	0

- **FF** e **FS**: numero totale di falli fatti e subiti
- **PS** e **PR**: numero di palle perse e recuperate
- **PUN**: numero di punti effettuati
- **RT**, **RO** e **RD**: rispettivamente il numero di rimbalzi totali, offensivi e difensivi
- **SD** e **SS**: stoppate date e subite
- **TL**, **T2** e **T3**: numero totale di tiri liberi, da due punti e da tre punti tentati
- **TLdone**, **T2done** e **T3done**: numero totale di tiri liberi, da due punti e da tre punti realizzati
- **TLmiss**, **T2miss** e **T3miss**: numero totale di tiri liberi, da due punti e da tre punti sbagliati

Nella Tabella B.1 riportata nell'Appendice B vengono forniti tutti i valori degli indicatori appena descritti per alcuni giocatori.

Utilizzando questi indici calcolati usando i dati relativi a tutte le partite disputate nel corso della *Regular Season*, sono stati poi calcolati altri indicatori che combinano quest'ultimi per fornire delle valutazioni globali dei giocatori e che consentono di dare una prima impressione su quali giocatori si sono distinti per le loro performance.

Il primo indice sintetico presentato è quello introdotto da Kai (Kay, 1966), denominato **TBPS** (*Total Basketball Proficiency Score*). Questo indice sebbene abbia subito delle modifiche vista l'introduzione del tiro da tre punti viene tuttora utilizzato per valutare i giocatori. La sua formula è

$$TBPS = 2 \times T2done - T2miss + TLdone - TLmiss + AS + RD + 2 \times RO + PR - FF - PP.$$

Negli anni successivi Gomez e Moll hanno proposto un nuovo indice denominato **IEG** (*Individual Efficiency at Games*, per il quale si veda ad esempio

Simovic *et al.*, 2019), che introduce nella formula anche le stoppate, e consente di valutare l'efficienza dei vari giocatori sia nel corso delle singole partite che nel corso di un'intera stagione. Questo viene calcolato come

$$IEG = PUN + (RT + PR + SD + FS) + 2 \times AS - (PP + T2miss + TLmiss).$$

Nel 1988, Dave Heeren (Heeren, 1988) introduce il **TENDEX**, indicatore che viene subito utilizzato da molti allenatori di NBA, e che sebbene sia stato poi parzialmente modificato, ha largo utilizzo ancora oggi. Questo viene calcolato come

$$TENDEX = [PUN + RT + AS + SD + PR - PP - (T2miss + T3miss) - 0.5 \times TLmiss - FF] / MIN.$$

Successivamente, dal 2003 in poi, vista l'evoluzione dei sistemi di raccolta dei dati, sono stati introdotti altri indici che integrano i precedenti considerando anche altri aspetti, come ad esempio le violazioni e le statistiche di squadra. Uno di questi è quello proposto dal Mays Consulting Group (per il quale si veda ad esempio Oliver, 2004, Paragrafo 5), che viene calcolato come

$$MM = (1.8 \times T2done) + (0.9 \times TLdone) + (3 \times T3done) + (0,65 \times RT) + (0.9 \times AS) + (0.8 \times SD) + PR - [0.65 \times (T2miss + T3miss)] - (0,5 \times TLmiss) - PP.$$

Infine, l'ultimo indice presentato nelle analisi è quello che viene utilizzato in NBA per definire l'**MVP** (Most Valuable Player), ovvero un indicatore realizzato dal centro di ricerca IBM (IBM, 1945). Questo è ad oggi uno degli indici più utilizzati nel basket e consente di fornire delle valutazioni sui giocatori considerando anche l'impatto che questi hanno all'interno della squadra. Quest'ultimo viene calcolato come

$$MVPIBM = [PUN + RT + AS + PR + SD - (T2 + T3) + FF + PP] + (partitevinte \times 10) \times 250 / [(tPUN + tRT + tAs + tPR + tSD - t(T2 + T3) - tFF - tPP)].$$

dove la t davanti alle sigle indica la somma per tutti i giocatori della squadra.

Tabella 1.4: Principali indici sintetici utilizzati nel Basket

Giocatore	Squadra	TBPS	IEG	TENDEX	MM	MVPIBM
Giannis Antetokounmpo	MIL	2403	3849	7.30	2874.25	249.26
Karl-Anthony Towns	MIN	1965	3257	5.64	2725.30	222.84
Kawhi Leonard	TOR	1575	2684	5.65	1840.51	209.96
Marc Gasol	TOR	1266	2476	4.63	1806.32	203.55
Pascal Siakam	TOR	1456	2417	5.49	1853.56	197.14
Kyle Lowry	TOR	1082	2644	5.43	1441.53	184.72
Andre Drummond	DET	2171	2647	6.93	2658.45	177.64
Tobias Harris	LAC	1445	2533	4.89	2087.26	171.05
Serge Ibaka	TOR	1214	1818	4.68	1618.09	170.98
Montrezl Harrell	LAC	1430	2180	6.16	1819.54	164.12

Nella Tabella 1.4 vengono presentati i risultati degli indici appena descritti calcolati per i 10 giocatori che presentano il valore più alto dell'indice MVPIBM (ordinati in modo decrescente per quest'ultimo).

Dai risultati presenti nella tabella emerge inizialmente che i giocatori più pericolosi e più in forma della stagione 2018/2019 sono Giannis Antetokounmpo vincitore sia del titolo di MVP della *Regular Season* che di miglior giocatore difensivo della stagione, Kawhi Leonard, anche vincitore del titolo di MVP delle *finals* e Karl-Anthony Towns, che nonostante le difficoltà riscontrate dalla propria squadra è riuscito a disputare una buona stagione. Un'altra cosa interessante che si evince dalla tabella, è il fatto che ben 5 dei 10 giocatori riportati appartenessero in quell'anno ai *Toronto Raptors*, ad indicare quindi una gran forma dei giocatori di questa squadra. Nella tabella si può comunque notare che i valori degli indici non seguono lo stesso ordine, e la valutazione dei giocatori cambia dunque a seconda dell'indicatore utilizzato. Per questo motivo, sebbene questi indici uniscano molte informazioni e consentano di fornire delle prime valutazioni sui giocatori, risulta particolarmente difficile valutare i cestisti selezionando uno di questi indicatori rispetto agli altri, e nasce quindi la necessità di introdurre delle valutazioni dei giocatori che coinvolgano l'uso di metodi probabilistici ed inferenziali.

I dati

Capitolo 2

Modello logistico per la valutazione dei giocatori NBA

Dopo quanto detto nel capitolo precedente, sembra sensato tentare di valutare le performance offensive dei giocatori utilizzando dei modelli statistici. In questo capitolo verrà quindi affrontato il problema oggetto di studio utilizzando delle tecniche inferenziali. In particolare, tramite i dati descritti nel capitolo precedente, si andrà ad analizzare la relazione tra la variabile risposta e le potenziali variabili esplicative, con il duplice scopo di studiare quali sono gli eventi che caratterizzano particolarmente un'azione pericolosa dal punto di vista offensivo e di prevedere la probabilità con cui una data azione viene considerata importante, ovvero la probabilità che questa porti alla realizzazione di un canestro. Sulla base dell'applicazione di questo modello e delle probabilità da esso stimate, verrà poi calcolato un *ranking* dei giocatori per stabilire quali siano i cestisti più efficaci per la stagione considerata. Tutti i calcoli e le analisi sui giocatori vengono svolte dopo ogni partita disputata, consentendo di effettuare delle valutazioni sia nel corso della stagione che alla fine della stessa, mettendo quindi a disposizione una serie storica delle prestazioni dei giocatori che permetterà di fornire valutazioni in tempo reale sul loro andamento.

2.1 Modello di regressione logistica

Il primo modello che è stato considerato, vista la natura dicotomica della variabile risposta, è un modello di regressione logistica, appartenente alla classe dei modelli lineari generalizzati (McCullagh e Nelder, 1989).

In particolare, indicate con y_1, y_2, \dots, y_n le osservazioni relative alla variabile risposta, si assume che queste siano realizzazioni indipendenti delle variabili casuali Y_1, Y_2, \dots, Y_n , con $Y_i \sim \text{Bernoulli}(\pi_i)$, e con π_i parametro ignoto compreso tra 0 e 1.

Successivamente, si pone in relazione la probabilità π_i con un insieme di variabili esplicative, definite come $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})$ per $i = 1, \dots, n$, attraverso un vettore di p parametri di regressione $\beta = (\beta_1, \dots, \beta_p)$. Più in dettaglio si definisce il predittore lineare η_i e lo si pone in relazione alla probabilità π_i . In particolare si assume che

$$g(\pi_i) = \eta_i = \mathbf{x}_i \beta = \beta_1 x_{i,1} + \dots + \beta_p x_{i,p},$$

con $g(\cdot)$ funzione di legame logistica, ossia

$$g(\pi_i) = \text{logit}(\pi_i) = \log \left(\frac{\pi_i}{1 - \pi_i} \right).$$

Questo implica

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}.$$

Il modello appena specificato porta alla seguente funzione di verosimiglianza, attraverso la quale verranno poi stimati i parametri β

$$L(\beta; X, \mathbf{y}) \propto \prod_{i=1}^n \left[\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \right]^{y_i} \times [1 + \exp(\eta_i)]^{y_i - 1},$$

dove $\mathbf{y} = y_1, \dots, y_n$ è il vettore relativo alla variabile risposta e X è la matrice del disegno, con i -esima riga X_i .

A questo punto non resta che definire le variabili esplicative che il modello descritto dovrà contenere. Inizialmente sono state quindi inserite tutte e sole le variabili che descrivono la tipologia di azione effettuata considerando anche l'esito di un'eventuale tiro a canestro, ma tralasciando l'informazione

relativa ai giocatori. In questo modo però si valutano tutti i giocatori allo stesso modo, senza tenere conto del ruolo del giocatore. Di conseguenza, per non perdere l'informazione relativa alla posizione, e per fornire delle valutazioni realistiche di quanto un giocatore possa incidere nel gioco data la sua posizione in campo, sono stati introdotti nel modello degli effetti fissi a rappresentare il ruolo del giocatore, permettendo di migliorare le capacità del modello inserendo un numero ristretto di parametri, portando a doverne stimare un totale di 58.

Per affrontare il problema di valutare le performance offensive dei cestisti sembra tuttavia sensato introdurre anche altri effetti fissi, ovvero quelli relativi a ciascun giocatore. In questo modo si può considerare non solo la posizione che i giocatori ricoprono quando sono sul terreno di gioco, ma si forniscono anche informazioni sul loro stato di forma in una specifica giornata. Questo consente anche di valutare quanto ciascun giocatore, a prescindere dall'azione che effettua, possa essere considerato pericoloso dal punto di vista offensivo. L'utilizzo di effetti fissi permette quindi di modellare l'eterogeneità tra soggetti diversi e allo stesso tempo ha il vantaggio di evitare l'endogeneità che si potrebbe introdurre utilizzando degli effetti casuali, cosa che porterebbe ad un aumento della distorsione nella stima dei parametri. Gli effetti fissi consentono inoltre di spiegare una parte della variabilità del processo osservato che altrimenti risulterebbe essere latente, portando quindi ad evitare degli errori sistematici nella stima.

Nel modello presentato in precedenza, l'introduzione degli effetti fissi porta quindi a calcolare la probabilità che un'azione i , con $i = 1, \dots, n$, svolta da un giocatore j , con $j = 1, \dots, J$ dove J è il numero totale di giocatori, sia rilevante dal punto di vista offensivo come

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})},$$

modificando il predittore lineare introducendo un termine α per ogni giocatore:

$$\eta_{ij} = \alpha_j + \mathbf{x}_i\beta.$$

L'introduzione di questi effetti porta quindi ad un numero molto elevato di parametri da stimare. Nel *dataset* utilizzato infatti, dopo aver svolto le

Tabella 2.1: Test del rapporto di verosimiglianza tra i modelli con e senza effetti fissi

Modello	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Senza effetti fissi	663898	727709			
Con effetti fissi	663447	726910	451	798.38	< 2.2e-16 ***

operazioni di pulizia rimangono ben 452 giocatori che corrispondono ad altrettanti effetti fissi, portando a dover stimare un modello con un numero di parametri pari a 509. La mole di dati disponibili in termini di numerosità di osservazioni sembra comunque permettere l’inserimento degli effetti fissi per i giocatori. Tuttavia, per giustificare questa operazione da un punto di vista statistico oltre che da un punto di vista logico, è stato effettuato un test del rapporto di verosimiglianza per confrontare il modello logistico che contiene gli effetti fissi per i giocatori e il modello che non li contiene. Nello specifico, la statistica test utilizzata per confrontare i modelli viene calcolata come:

$$D = \text{Devianza}_{SE} - \text{Devianza}_{FE}$$

Con Devianza_{SE} che corrisponde alla devianza del modello senza effetti fissi e con Devianza_{FE} che rappresenta quella relativa al modello che li contiene. Questa è stata poi confrontata con la distribuzione χ^2 con un numero di gradi di libertà pari alla differenza dei parametri tra il modello completo e quello ridotto. Per effettuare questo test, i modelli sono stati stimati sulla totalità dei dati a disposizione, e ciò ha portato ad ottenere i risultati descritti in Tabella 2.1.

Da quest’ultima, si può notare come il p -value ottenuto dal test del rapporto di verosimiglianza sia particolarmente basso, portando a considerare l’introduzione di questi effetti come significativa nella valutazione dei giocatori.

In Tabella 2.2 vengono quindi riportati tutti i coefficienti stimati con questo modello relativi a tutti gli eventi e ai sotto-eventi appartenenti alla categoria tiri, mentre in Tabella 2.3 vengono riportati i coefficienti stimati per i rimanenti eventi e sotto-eventi. Per questioni di spazio, non vengono qui

riportate le stime relative agli effetti fissi dei giocatori, che vengono tuttavia riportate nell'Appendice B in Tabella B.2.

Andando inizialmente a valutare i coefficienti relativi ai tiri stimati dal modello, si nota che il coefficiente relativo al tiro nella sua categorizzazione più generale ha il segno negativo, portando a dire che tirare a canestro porta ad una diminuzione della probabilità di far risultare pericolosa un'azione dal punto di vista offensivo. Questo aspetto non sembra avere senso dal punto di vista logico, ma può essere imputato al fatto che non tutti i tiri si concludono con la realizzazione del canestro, portando talvolta anche a delle pericolose ripartenze da parte della squadra avversaria. Valutando invece i sotto-eventi ad esso associati, è interessante notare che i tiri che portano più probabilmente a rendere pericolosa un'azione sono quelli identificati come "Layup" e "Running Layup", due tipologie di tiri da due punti che sono effettuati molto vicino al canestro. Per quanto riguarda invece le altre tipologie di eventi, si nota che commettere un fallo o perdere palla porta il modello a giudicare negativamente l'azione, mentre recuperare palla, fare un assist o realizzare un canestro, come logicamente ci si aspetta, hanno un impatto positivo nella valutazione dell'azione a livello offensivo, risultando particolarmente significativi.

Analizzando invece il modello dal punto di vista della corretta classificazione delle azioni si ottengono i risultati riportati in Tabella 2.4, dove nelle colonne vi sono i valori predetti dal modello, mentre nelle righe vi sono i veri valori della variabile risposta. Per il calcolo delle previsioni, sono stati assegnati valori uguali ad 1 se la probabilità stimata dal modello risulta maggiore o uguale a 0.5 e valori uguali a 0 altrimenti.

Tabella 2.4: Tabella di corretta classificazione

Valori osservati	Valori predetti	
	0	1
0	356049	85549
1	105854	116504

Tabella 2.2: Coefficienti stimati nel modello di regressione logistica con effetti fissi per gli eventi e i sotto-eventi relativi ai tiri

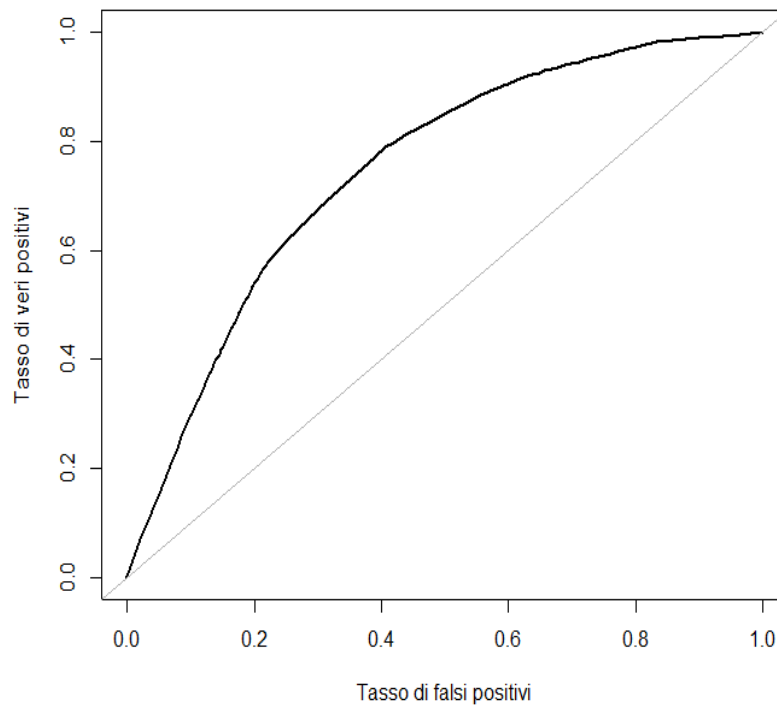
Variabile	Stima	Standard Error	z value	Pr(> z)
shot	-0.3135	0.0152	-20.5855	3.7015e-94
Alley_Oop_Dunk	-0.1826	0.0482	-3.7848	1.5382e-04
Alley_Oop_Layup	0.0020	0.0660	0.0307	9.7547e-01
Driving_Bank_Hook_Shot	-0.2558	0.1579	-1.6198	1.0526e-01
Driving_Finger_Roll_Layup	-0.0561	0.0339	-1.6531	9.8298e-02
Driving_Hook_Shot	-0.2225	0.0605	-3.6794	2.3379e-04
Driving_Reverse_Layup	-0.1128	0.0472	-2.3899	1.6848e-02
Dunk	0.0043	0.0433	0.1005	9.1994e-01
Fadeaway_Jumper	-0.1735	0.0390	-4.4487	8.6364e-06
Finger_Roll_Layup	-0.0014	0.0928	-0.0151	9.8791e-01
Floating_Jump_Shot	-0.0767	0.0339	-2.2609	2.3765e-02
Hook_Bank_Shot	-0.3451	0.1528	-2.2586	2.3907e-02
Hook_Shot	-0.1418	0.0395	-3.5910	3.2936e-04
Jump_Bank_Shot	-0.2038	0.0733	-2.7811	5.4168e-03
Jump_Shot	-0.1997	0.0132	-15.0837	2.0718e-51
Layup	0.0775	0.0232	3.3405	8.3633e-04
Pullup_Jump_Shot	-0.2072	0.0181	-11.4786	1.6899e-30
Putback_Dunk	-0.0121	0.0895	-0.1348	8.9277e-01
Putback_Layup	0.0606	0.0414	1.4642	1.4312e-01
Reverse_Dunk	-0.1357	0.1875	-0.7236	4.6930e-01
Reverse_Layup	0.02995	0.0524	0.5711	5.6793e-01
Running_Dunk	-0.0020	0.0487	-0.0420	9.6645e-01
Running_Finger_Roll_Layup	-0.0010	0.0709	-0.0146	9.8831e-01
Running_Jump_Shot	0.0257	0.0604	0.4249	6.7084e-01
Running_Layup	0.1709	0.0326	5.2433	1.5774e-07
Running_Reverse_Layup	0.0617	0.1129	0.5462	5.8495e-01
Step_Back_Jump_Shot	-0.2776	0.0282	-9.8518	6.7320e-23
Turnaround_Bank_Hook_Shot	-0.1886	0.1408	-1.3394	1.8045e-01
Turnaround_Fadeaway_	-0.2394	0.0463	-5.1693	2.3493e-07
Turnaround_Hook_Shot	-0.2874	0.0448	-6.4142	1.4152e-10
Turnaround_Jump_Shot	-0.1835	0.0417	-4.3975	1.0946e-05

2.1 Modello di regressione logistica

Tabella 2.3: Coefficienti stimati nel modello di regressione logistica con effetti fissi relativi a tutti gli eventi e i sotto-eventi ad eccezione dei tiri

Variabile	Stima	Standard Error	z value	Pr(> z)
foul	-1.6729	0.0214	-78.3595	0.0000e+00
free_throw	-0.1056	0.0152	-6.9531	3.5726e-12
jump_ball	0.2079	0.0499	4.1919	2.7657e-05
turnover	-0.6974	0.0324	-21.5428	6.1804e-103
violation	-0.1571	0.0604	-2.5979	9.3806e-03
backcourt	-0.2378	0.2327	-1.0219	3.0683e-01
bad_pass_	-1.0927	0.0516	-21.1826	1.3823e-99
discontinue_dribble_turnover	0.4343	0.2977	1.4588	1.4462e-01
double_dribble_turnover	0.1778	0.2915	0.6099	5.4187e-01
lost_ball_	-1.0605	0.0640	-16.5622	1.3066e-61
offensive_goaltending	0.3381	0.2134	1.5843	1.1313e-01
out_of_bounds_lost_ball_turnover	-0.0537	0.0755	-0.7112	4.7694e-01
palming_turnover	0.0179	0.3590	0.0499	9.6013e-01
rebound_defensive	0.6269	0.0119	52.3299	0.0000e+00
rebound_offensive	0.7810	0.0159	48.9294	0.0000e+00
step_out_of_bounds_turnover	0.1188	0.1071	1.1098	2.6707e-01
traveling_	0.0449	0.0685	0.6557	5.1204e-01
assist	1.6219	0.0128	127.1717	0.0000e+00
steal	0.8985	0.0174	51.5840	0.0000e+00
block	0.2568	0.0221	11.6338	2.7744e-31
res	1.7923	0.0093	193.5227	0.0000e+00

Figura 2.1: Curva ROC ottenuta con il modello di regressione logistico



Da quest'ultima si ottiene che il tasso di corretta classificazione è pari a 0.712, mentre si ha una percentuale di falsi negativi uguale al 23% ed una percentuale abbastanza elevata di falsi positivi, pari al 42%, portando ad affermare che il modello tende a selezionare come importanti dal punto di vista offensivo azioni che in realtà non lo sono. Quello che si vuole prevedere non è tuttavia il vero valore della variabile risposta, poiché questo non descrive effettivamente il vero oggetto di interesse, ma rappresenta una sua *proxy*. Di conseguenza, anche se il modello non sembra classificare le azioni con una precisione particolarmente elevata, i risultati vengono considerati accettabili. Una migliore valutazione della capacità predittiva del modello si ottiene con la curva ROC (si veda ad esempio Salvan *et al.*, 2020, Paragrafo 3.7). Il grafico della curva ROC mostrato in Figura 2.1 porta ad ottenere un'area sotto la curva (AUC) pari a 0.748, risultando quindi accettabile.

2.2 Problemi del modello e correzione di Firth

Una cosa particolarmente importante di cui tenere conto, riguarda però il fatto che il modello descritto in precedenza non verrà stimato su tutti i dati, ma verrà stimato dopo ogni giornata in cui è stata disputata almeno una partita, utilizzando quindi un numero di osservazioni molto inferiore rispetto alla totalità delle osservazioni disponibili. Questo aspetto, vista l'elevata quantità di parametri da stimare, introduce quindi un problema rilevante, dovuto al fatto che le proprietà dello stimatore di massima verosimiglianza deteriorano quando il numero di parametri è elevato e si avvicina al numero di osservazioni. Nella stima dei modelli adattati dopo ciascuna giornata di campionato infatti, non si ottengono dei risultati particolarmente buoni, riscontrando anche dei problemi di convergenza, che portano sia ad una stima non corretta dei parametri che anche alla previsione di probabilità di pericolosità delle azioni a livello offensivo che sono numericamente uguali a 0 o ad 1. In particolare, andando a valutare di volta in volta le stime ottenute dai modelli, si nota che si ottengono sia coefficienti che deviazioni standard con valori molto elevati, facendo emergere il problema relativo alle stime infinite (si veda ad esempio Salvan *et al.*, 2020, Paragrafo 3.4.2).

Queste complicazioni, sono principalmente dovute alla numerosità dei parametri, ma più formalmente sorgono per via della quasi perfetta separazione tra le variabili, ovvero poiché per determinate configurazioni di variabili ci sono solamente giocatori che appartengono ad una determinata classe della variabile risposta, portando così a sovradattare i dati. Per questo motivo si ha bisogno di un metodo di stima per dati binari che risolva questi problemi, e nel seguito verrà quindi utilizzato un nuovo metodo di stima dei parametri, il quale introduce una correzione per la distorsione, che verrà poi applicato al presente caso di studio.

Prima di passare all'implementazione del metodo, viene quindi presentata brevemente la correzione per la distorsione proposta da Firth (Firth, 1993) per la stima dei parametri. In sintesi, quello che viene fatto con questo metodo corrisponde ad effettuare una modifica della funzione punteggio volta a ridurre la distorsione delle stime di massima verosimiglianza.

Formalmente, si ha che in un problema regolare di stima, le stime di

massima verosimiglianza vengono ottenute come soluzione delle equazioni di verosimiglianza. In particolare, indicando con $L(\beta)$ la funzione di verosimiglianza e con $l(\beta)$ la funzione di log-verosimiglianza, le equazioni utilizzate per la stima dei parametri che portano ad ottenere $\hat{\beta}$ sono date da:

$$\nabla l(\beta) = U(\beta) = 0$$

Firth (Firth, 1993) propone un metodo per ridurre la distorsione delle stime $\hat{\beta}$ così ottenute introducendo una piccola distorsione nella funzione punteggio $\nabla l(\beta)$. Più precisamente, se si ha che $\hat{\beta}$ presenta una distorsione pari a $b(\beta)$, la funzione punteggio viene spostata in basso in ogni punto β di una quantità pari a $i(\beta)b(\beta)$, con $i(\beta) = E[-\nabla U(\beta)]$ che corrisponde alla matrice nota come informazione di Fisher. Ciò porta a definire la nuova funzione punteggio

$$U^*(\beta) = U(\beta) - i(\beta)b(\beta),$$

la quale porta a calcolare $\hat{\beta}$, la cui distorsione sarà minore rispetto a quella delle stime ottenute con le usuali equazioni, come soluzione di:

$$U^*(\beta) = 0.$$

Venendo ora al problema oggetto di studio, si ha che il parametro da stimare corrisponde ad un parametro di un modello appartenente ad una famiglia esponenziale con parametro canonico β (Pace e Salvan, 1996). Per questi modelli, stimare i parametri utilizzando la correzione per la distorsione proposta, corrisponde a massimizzare la seguente verosimiglianza penalizzata

$$L^*(\beta) = L(\beta)|i(\beta)|^{\frac{1}{2}}$$

Per quanto riguarda quest'ultima equazione, è particolarmente interessante notare che $|i(\beta)|^{\frac{1}{2}}$ corrisponde alla distribuzione a priori di Jeffreys (Kosmidis e Firth, 2020), facendo sì che in questa tipologia di modelli la distorsione venga corretta attraverso la moda di una distribuzione a posteriori basata sulla a priori precedentemente citata.

Con l'applicazione di questo metodo, i parametri stimati in un modello di regressione logistica subiscono un effetto di *shrinkage* verso lo zero, e questo può essere spiegato poiché, data la matrice del disegno indicata con X , si ha

che la matrice d'informazione è data da $i(\beta) = X^T W X$, con $W = \text{diag}(\pi_i(1 - \pi_i))$, il cui determinante è massimizzato quando $\pi_i = 0.5$, ovvero quando $\beta = \mathbf{0}$, facendo sì che la distribuzione a priori di Jeffreys comprima le stime appunto verso questo valore.

Per la stima del modello di regressione logistica con correzione di Firth è stato utilizzato il pacchetto R `brglm2` (Kosmidis *et al.*, 2020) il quale permette appunto di stimare dei modelli lineari generalizzati utilizzando la correzione per la distorsione presentata.

2.3 Costruzione del *ranking*

Dopo aver presentato il modello utilizzato per la stima, viene ora esposto il metodo con il quale verranno valutati i giocatori. Lo scopo principale è quello di utilizzare il modello stimato in precedenza per valutare le performance offensive dei giocatori aggiornandole dopo ogni giornata in cui è stata disputata almeno una partita, così da fornire delle valutazioni in tempo reale dei giocatori. Il *ranking* verrà calcolato sulla base del modello contenente tutte le variabili introdotte nel primo capitolo, includendo anche quella che indica se il giocatore ha realizzato un eventuale tiro a canestro, in modo tale da valutare l'azione nel suo complesso, comprendendo anche la capacità di finalizzazione, aspetto che nel basket è fondamentale visto l'ammontare di punti segnati in ciascuna partita.

Lo scopo non è tuttavia fornire delle valutazioni isolate, contemplando solamente le performance del giocatore in ogni singola partita, poiché così non si andrebbe a considerare l'importanza della continuità del giocatore e della sua evoluzione nel corso della stagione, perdendo così un aspetto molto rilevante. L'obiettivo è perciò quello di valutare ogni giocatore esaminando non solo la partita appena disputata ma includendo nella valutazione anche l'informazione relativa alle partite precedenti, permettendo così delle considerazioni più complete ed esaustive.

2.3.1 Primi metodi e problemi computazionali

Il primo metodo che è stato implementato per valutare le performance dei giocatori è stato stimare il modello di regressione logistica con correzione di Firth di giornata in giornata, utilizzando ogni volta tutti i dati disponibili dall'inizio della stagione fino all'istante temporale considerato. Così facendo però sorgono due problemi. Il primo corrisponde al fatto di pesare allo stesso modo tutte le partite considerate, non consentendo di valutare lo stato di forma dei giocatori in modo dinamico. Il secondo problema è invece di natura computazionale.

Per quanto riguarda il primo problema, questo può essere risolto con l'introduzione dei pesi, ovvero pesando opportunamente le osservazioni a seconda della partita alla quale si riferiscono. Ciò implica tuttavia il dover scegliere i pesi da applicare, cosa che non è immediata, e che potrebbe necessitare la consultazione di esperti nell'ambito della valutazione dei giocatori di pallacanestro.

Il secondo problema è invece più complicato. Quando si è nel caso di dati che arrivano a flussi infatti, ovvero quando si lavora con i cosiddetti *data stream*, l'aggiornamento dei parametri del modello diventa particolarmente complicato, poiché per ogni nuova osservazione, o come in questo caso per ogni nuova partita disponibile, il problema richiede di ristimare da capo il modello, facendo aumentare il tempo d'attesa per ottenere i risultati. Nel presente caso si ha inoltre che la variabile risposta è dicotomica e che il modello che viene utilizzato appartiene alla classe dei modelli lineari generalizzati. Contrariamente al caso dei modelli lineari infatti, dove è possibile aggiornare semplicemente le stime dei parametri con delle tecniche derivanti dall'algebra lineare senza dover ristimare ogni volta il modello (Azzalini e Scarpa, 2004), nel caso dei modelli lineari generalizzati non esistono delle tecniche esatte per effettuare l'aggiornamento dei parametri ogni qual volta una nuova osservazione, o un blocco di osservazioni, diventa disponibile.

Per tentare di risolvere il problema, sono state applicate delle tecniche di discesa stocastica del gradiente. Con questi metodi, per ogni nuovo gruppo di osservazioni che diventa disponibile, è quindi possibile aggiornare le stime dei parametri utilizzando appunto il gradiente della funzione di

log-verosimiglianza, ottenendo così delle approssimazioni delle vere stime di massima verosimiglianza. Per l'implementazione di questo metodo è stato utilizzato il pacchetto R *sgd* (Kim *et al.*, 2019), che permette di utilizzare la discesa stocastica del gradiente nell'ambito dei modelli lineari generalizzati. I dati a disposizione però presentano il problema della quasi perfetta separazione, e quindi l'utilizzo di questo metodo non porta a buoni risultati, poiché per essere utilizzato bisogna che le colonne della matrice del disegno siano indipendenti. Per questo motivo, nel caso in esame le stime che si ottengono sono molto diverse dalle stime dei parametri ottenute ristimando il modello di volta in volta, portando a scartare questa opzione.

2.3.2 La statistica EWMA per aggregare i risultati

Per i problemi elencati in precedenza è stato deciso di percorrere un'altra strada, in modo tale da risolvere sia i problemi di natura computazionale, sia i problemi relativi al fatto di considerare nella valutazione del giocatore anche le partite precedenti a quella disputata in modo dinamico. L'idea di base di questo metodo è quella di stimare un modello utilizzando solamente i dati che diventano disponibili dopo ogni giornata di campionato, e successivamente aggregare i risultati ottenuti utilizzando un metodo statistico opportuno.

Più precisamente, al termine di ogni giornata di campionato nella quale si disputa almeno una partita, verrà stimato un modello utilizzando solamente i dati relativi a quella giornata, e dopodiché si procederà prima valutando le performance del giocatore relativamente alla giornata considerata, e successivamente si andranno ad aggregare i risultati per considerare opportunamente anche le giornate precedenti.

Una cosa importante di cui tenere conto prima di procedere con le analisi è data dal fatto che alcuni giocatori presenti all'interno del *dataset* potrebbero giocare poco nel corso di alcune partite. Di conseguenza, per fornire delle valutazioni sensate e per far sì che non ci siano problemi nella stima degli effetti fissi dei cestisti, verranno valutati in ciascuna giornata solamente i giocatori che hanno effettuato almeno 5 tocchi di palla.

A questo punto, al termine di ogni giornata in cui si sono giocate delle partite, si utilizzeranno i nuovi dati disponibili e si adatterà a quest'ultimi

il modello di regressione logistica con correzione di Firth presentato nella sezione precedente, il quale conterrà appunto tutte le variabili relative alle azioni e tutti gli effetti fissi, sia quelli riguardanti il ruolo che quelli relativi ai cestisti che hanno effettuato almeno 5 giocate nella giornata considerata. Successivamente, utilizzando il modello appena stimato per ciascuna giornata t , con $t = 1, \dots, T$, dove T indica il numero totale di giornate in cui è stata disputata almeno una partita, si andrà a stimare per ogni azione svolta la probabilità che questa sia pericolosa dal punto di vista offensivo, ottenendo il vettore: $\hat{\pi}_t = (\hat{\pi}_{1,t}, \hat{\pi}_{2,t}, \dots, \hat{\pi}_{k,t})$, dove con k viene indicato il numero di azioni che si sono svolte nel corso della giornata t -esima. Dopodiché si andranno ad aggregare le probabilità per ciascun giocatore separatamente, facendo una media di tutte le probabilità di incisione nella realizzazione di un canestro stimate dal modello. Quindi, indicando con J il numero totale di giocatori scesi in campo nella giornata t , si utilizzerà la formula (2.1) per fornire una valutazione per ciascun giocatore r nella giornata t , con $r = 1, \dots, J$, dove con $\hat{\pi}_{i,t,r}$ vengono indicate le probabilità calcolate in precedenza e relative al giocatore r nella giornata t nell'azione i , con $n_{r,t}$ il numero di azioni alle quali il giocatore r ha partecipato nella giornata t e con n_t il numero di osservazioni relativo alla giornata t . In dettaglio, si ha che:

$$\hat{\pi}_{i,t,r} = \begin{cases} 0, & \text{se } \hat{\pi}_{i,t} \text{ non è associata al giocatore } r, \\ \hat{\pi}_{i,t}, & \text{altrimenti,} \end{cases}$$

e si ottiene quindi la formula per la valutazione di ciascun giocatore in una data giornata

$$\widehat{\text{Player}}_{r,t} = \frac{1}{n_{r,t}} \sum_{i=1}^{n_t} \hat{\pi}_{i,t,r} \quad (2.1)$$

In questo modo si ottengono una serie di valori compresi tra 0 e 1 che permettono di valutare le performance di ciascun giocatore nella partita considerata.

Come già detto in precedenza però, l'obiettivo non è quello di fornire delle valutazioni isolate, bensì è quello di valutare i giocatori durante tutta la stagione, utilizzando di volta in volta anche tutta l'informazione relativa alle partite precedenti a quella considerata in un determinato istante di tempo.

Per aggregare i risultati delle varie partite disputate, un metodo che viene ampiamente utilizzato per aggregare i risultati in applicazioni statistiche che riguardano la sorveglianza, è quello che prevede l'utilizzo della statistica EWMA (*Exponentially Weighted Moving Average*) (Hunter, 1986). In particolare, facendo ricorso a questa statistica si possono aggregare i risultati consentendo di utilizzare nella valutazione dei giocatori per una data partita anche tutta l'informazione relativa alle giornate precedenti. Questa statistica viene definita nel presente caso di studio come:

$$\widehat{\text{Rank}}_{r,t} = \begin{cases} \widehat{\text{Player}}_{r,t}, & \text{se } t = 1, \\ (1 - \lambda)\widehat{\text{Rank}}_{r,t-1} + \lambda\widehat{\text{Player}}_{r,t}, & \text{altrimenti,} \end{cases} \quad (2.2)$$

dove λ è una costante fissata in $(0, 1)$.

Per quanto riguarda la statistica EWMA, questa viene utilizzata nell'ambito delle carte di controllo, dove l'obiettivo è quello di sorvegliare un processo considerando anche l'evoluzione nel tempo del processo stesso. Tuttavia, i suoi ambiti applicativi sono molteplici e vanno dalla biomedicina alla produzione industriale, fino ad arrivare come nel caso oggetto di studio in questa tesi a quello sportivo.

Nella formula (2.2) viene utilizzata nel calcolo una costante λ , la quale riveste un ruolo cruciale nella statistica presentata, permettendo di definire quanto sarà influente la partita considerata rispetto a tutte le partite precedenti. Per la sua scelta bisogna considerare che utilizzando un valore piccolo di λ si andrà a pesare poco la partita corrente e di più le partite precedenti; viceversa, utilizzando un valore grande di λ , si andrà a pesare molto la partita considerata dando meno peso allo storico delle partite. Nella Figura 2.2 viene riportata l'evoluzione delle valutazioni fornite per il giocatore LeBron James nel corso della stagione per diversi valori di λ , per mostrare appunto come questo influenzi in modo decisivo i risultati. I valori di λ utilizzati sono rispettivamente 0.3, 0.5, 0.7 e 0.9. Nel seguito, nel calcolo finale delle valutazioni dei giocatori è stato utilizzato un valore $\lambda = 0.5$, il quale consente di ottenere un buon compromesso e fornisce risultati soddisfacenti dal punto di vista qualitativo, permettendo di ottenere delle valutazioni sui cestisti per lo più simili a quelle proposte da esperti nel settore.

Figura 2.2: Evoluzione delle valutazioni di LeBron James per $\lambda \in \{0.3, 0.5, 0.7, 0.9\}$ (partendo dall'angolo in alto a sinistra in senso orario)

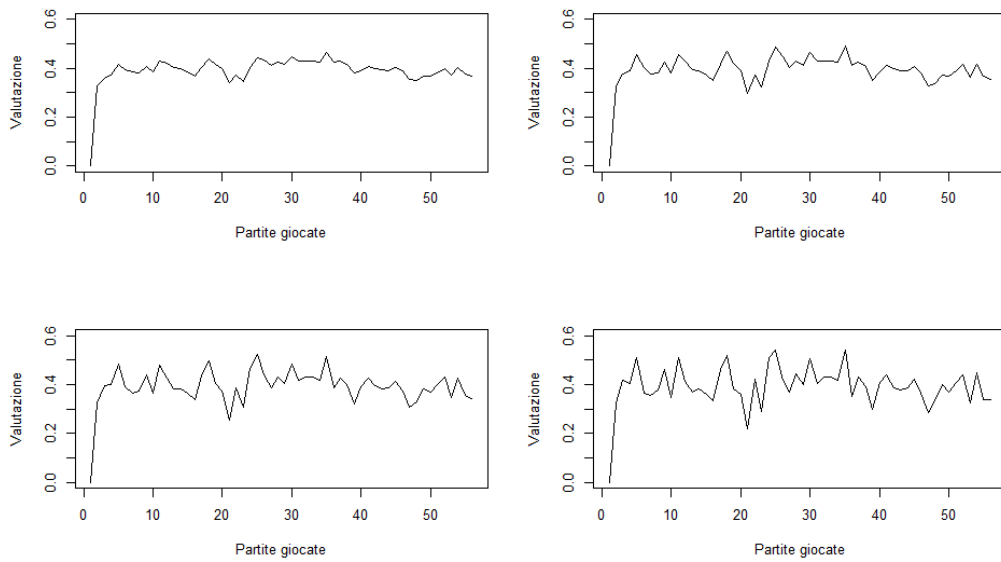
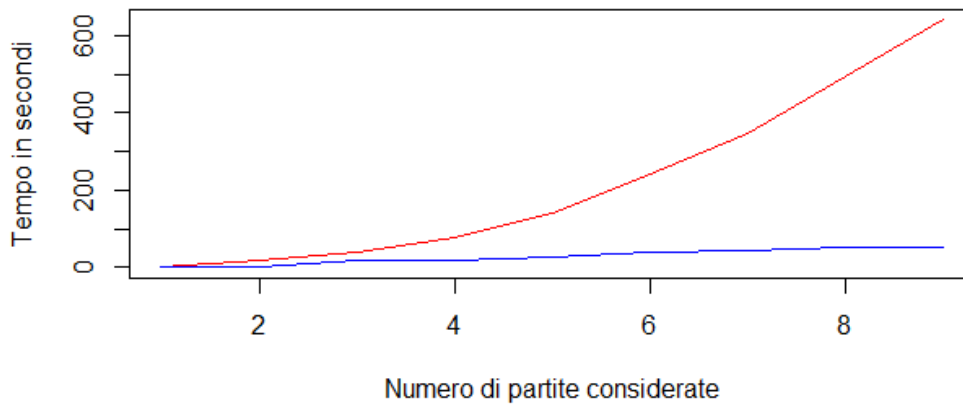


Figura 2.3: Evoluzione dei tempi computazionali con i due metodi presentati. La linea rossa si riferisce al primo metodo utilizzato, mentre la linea blu al metodo che utilizza la statistica EWMA



Con questo metodo si riesce dunque a fornire delle valutazioni partita dopo partita includendo tutte le informazioni disponibili. In termini di tempi computazionali, questo approccio è molto più conveniente rispetto al precedente, portandolo ad essere preferibile anche sotto questo punto di vista. Dal grafico riportato in Figura 2.3, che per questioni di leggibilità è stato costruito utilizzando le informazioni relative alle prime 8 giornate di campionato, si può notare infatti come per il primo metodo presentato (in rosso), dove per ogni partita è stato stimato un modello utilizzando tutti i dati dal primo match fino alla partita considerata, all'aumentare delle partite disponibili i tempi di esecuzione non aumentino linearmente ma esponenzialmente, portando brevemente col passare delle giornate a raggiungere tempi di esecuzione particolarmente lunghi e proibitivi. Per quanto riguarda invece il metodo che utilizza la statistica EWMA (in blu), si ha che il calcolo del *ranking* dopo ogni partita porta a dei tempi computazionali che evolvono linearmente, e che subiscono cambiamenti dovuti principalmente alla numerosità delle osservazioni considerate, facendolo risultare maggiormente interessante.

2.3.3 Calcolo del *ranking*

Arrivati a questo punto si hanno a disposizione le valutazioni di ogni giocatore partita dopo partita, portando a poter utilizzare per valutare i giocatori una serie storica contenente le analisi delle performance per ciascuno di essi, consentendo di studiarne quindi anche l'evoluzione nell'arco di tutta la stagione.

L'obiettivo finale della tesi è però quello di costruire una classifica dei giocatori globale, e una stratificata per ruolo, per capire chi si è comportato meglio dal punto di vista offensivo nel corso dell'anno, ed effettuare poi alcuni confronti sulla base dei risultati. Per fare ciò si può quindi per ogni giornata disponibile e per ogni giocatore, utilizzare le osservazioni relative alla giornata oggetto di studio della serie storica ottenuta con i metodi precedenti, avendo così a disposizione un valore sulla base del quale si potranno fare vari confronti tra i giocatori.

In questo modo però non si considera l'informazione relativa al numero di partite disputate e il rispettivo quantitativo di azioni effettuate, portando a delle valutazioni non propriamente ragionevoli, in quanto così facendo potrebbe essere che alcuni giocatori che effettuano poche azioni ma che in queste risultano particolarmente incisivi, vengano valutati in modo migliore rispetto ad altri giocatori che partecipando attivamente alla maggior parte delle azioni commettono inevitabilmente anche alcuni errori, perdendo quindi valore utilizzando il metodo in questione pur essendo pericolosi e fondamentali per la squadra.

Per sopperire a questo problema è stato quindi deciso di effettuare una correzione nelle valutazioni ottenute in precedenza, pesandole con il rapporto tra il numero di azioni e il numero di partite giocate fino alla giornata considerata. In particolare si andrà a calcolare il *ranking* corretto (per ogni giornata disponibile) come:

$$\widehat{\text{FinalRank}}_{r,t} = \frac{\widehat{\text{Rank}}_{r,t} \cdot \text{numAzioni}_{r,t}}{\text{numPartite}_{r,t}} \quad (2.3)$$

Dove $\text{numAzioni}_{r,t}$ e $\text{numPartite}_{r,t}$ con $r = 1, \dots, J$ indicano rispettivamente il numero totale di azioni e di partite alle quali il giocatore r partecipa fino alla giornata t considerata nel calcolo del *ranking*.

2.4 Risultati

Utilizzando la formula (2.3) si può quindi stilare una classifica dei giocatori sulla base delle stime delle loro performance ottenuti dall'applicazione del modello di regressione logistica ad effetti fissi con correzione di Firth. Questa classifica può essere sia globale, considerando tutti i giocatori congiuntamente e valutandoli indipendentemente dal ruolo che questi occupano, sia stratificata per ruolo, così da poter effettuare dei confronti più ragionevoli e mirati.

Nel seguito vengono presentati i risultati ottenuti distinguendo tra questi due casi, confrontando le valutazioni fornite dal modello sia con gli indici statistici descrittivi che con altre valutazioni fornite da esperti del settore. Per il calcolo del *ranking* si usano i risultati ottenuti dalla procedura descritta utilizzando tutti i dati fino all'ultima partita disputata, portando quindi ad avere dei risultati che riassumono l'andamento dei giocatori nel corso dell'intera stagione. Infine, per consentire uno studio più approfondito del problema, verranno utilizzate le valutazioni sui giocatori per fornire delle considerazioni anche sulle varie squadre partecipanti al campionato. Quest'ultima analisi viene fatta con lo scopo sia di capire quali squadre sono le squadre più orientate alla fase offensiva, che anche per capire dove giocano i cestisti più talentuosi e promettenti.

2.4.1 Valutazione globale dei giocatori

Per prima cosa vengono qui analizzati i risultati ottenuti a livello globale, senza fare distinzione tra i giocatori a livello di ruolo. Dall'applicazione del modello, si ottiene la classifica dei giocatori riportata in Tabella 2.5. In quest'ultima vengono visualizzati solamente i 10 giocatori ai quali viene attribuito il giudizio più alto, mentre in Tabella B.3 nell'Appendice B viene riportata la tabella con la classifica di tutti i giocatori disponibili. Per ciascuno dei cestisti presenti viene quindi indicato il ruolo, la squadra di appartenenza, il *ranking* calcolato senza la correzione per il rapporto tra minuti giocati e partite disputate e quello calcolato utilizzando la correzione presentata nella formula (2.2).

Tabella 2.5: Classifica dei 10 migliori giocatori NBA selezionati dal modello

Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
Giannis Antetokounmpo	MIL	PF	0.395	23.8
James Harden	HOU	PG	0.370	23.0
Anthony Davis	NOP	C	0.416	22.1
Nikola Jokic	DEN	C	0.398	20.2
Russell Westbrook	OKC	PG	0.322	19.6
Luka Doncic	DAL	SG	0.407	19.1
LeBron James	LAL	SF	0.353	19.1
Joel Embiid	PHI	C	0.311	18.5
Jrue Holiday	NOP	SG	0.395	17.6
Paul George	OKC	SF	0.342	17.4

Valutando ora i giocatori presenti in questa classifica, si nota che l'aggiustamento per il rapporto tra il numero di azioni e il numero di partite influisce molto nella classifica dei giocatori. Questa correzione permette di ottenere risultati più sensati e risulta particolarmente rilevante, indicando che non possa non essere considerata nel calcolo del *ranking*.

Per quanto riguarda i giocatori invece, si ha che al primo posto si trova Giannis Antetokounmpo, il quale nella stagione 2018/2019 ha svolto un ruolo fondamentale durante la *Eastern Conference*, aiutando la sua squadra a vincere ben 60 partite. Dopo aver ottenuto ottimi risultati durante la *Regular Season*, il greco viene anche nominato MVP della stagione venendo incluso sia nel primo quintetto *All-NBA* che nel primo quintetto *All-Defense*. Per queste motivazioni sembra quindi coerente che questo giocatore si trovi al primo posto, portando i risultati ottenuti dal modello statistico ad essere in accordo con il giudizio degli esperti nel settore.

Al secondo posto della classifica troviamo poi James Harden che con 36,1 punti di media a partita termina la stagione risultando il miglior marcatore della lega, arrivando secondo nella classifica che elegge l'MVP (Basketball-reference, 2019).

Per quanto riguarda invece il terzo e il sesto posto, questi sono occupati

da Anthony Davis e Luka Doncic, i quali si sono distinti nel corso dell'anno per vari aspetti. Il primo è stato uno dei migliori giocatori difensivi della stagione, mentre il secondo ha vinto il premio di *Rookie of the Year*, ovvero è stato considerato la migliore matricola del campionato, rendendolo così uno dei giocatori più appetibili per l'anno successivo.

Nelle posizioni rimanenti si trovano poi altri giocatori che hanno svolto un'ottima stagione, alcuni dei quali sono stati riconosciuti di vari premi come ad esempio l'inserimento nei migliori quintetti dei giocatori di NBA per la stagione considerata. Il modello e i metodi applicati a questo caso di studio sembrano quindi funzionare abbastanza bene, poiché riescono ad identificare i migliori giocatori del campionato, in accordo con i premi attribuiti dagli esperti.

Una cosa interessante da considerare riguarda il confronto con gli indici descrittivi presentati nel primo capitolo. Con il *ranking* appena calcolato, si arriva infatti ad escludere dai giocatori più pericolosi dal punto di vista offensivo Kawhi Leonard ed altri giocatori ai quali veniva assegnato un punteggio elevato nella formula che viene utilizzata per definire l'MVP (l'indice più utilizzato in NBA). Questo può essere dovuto al fatto che il metodo implementato non considera la bravura del giocatore rapportata alla squadra, come invece veniva fatto nel calcolo dell'MVP, ma analizza le sue performance separatamente e sembra quindi sensato che i risultati siano parzialmente diversi. In ogni caso, utilizzando la metodologia proposta in questo capitolo risulterebbe comunque semplice introdurre questo aspetto nel calcolo del *ranking*, poiché sarebbe sufficiente aggiungere al modello un effetto fisso relativo alla squadra. Questo aspetto, poiché l'interesse della tesi è quello di valutare le performance dei cestisti singolarmente, non è stato preso in considerazione, ma potrà essere oggetto di studi futuri.

Un ultimo aspetto che viene qui discusso relativo alla procedura statistica adottata, riguarda il fatto che tra i giocatori che vengono indicati come i migliori dal punto di vista offensivo in questo capitolo, sono in realtà presenti dei giocatori che svolgono compiti maggiormente orientati alla difesa, come ad esempio Giannis Antetokounmpo. Questo è dovuto al fatto che, essendo la pallacanestro uno sport dove l'azione si conclude molto rapidamente, le giocate difensive, i recuperi palla e le ripartenze assumono un ruolo fonda-

Tabella 2.6: Migliori 10 giocatori di ruolo ala grande classificati dal modello

Giocatore	Squadra	Ranking	Ranking corretto
Giannis Antetokounmpo	MIL	0.395	23.8
Blake Griffin	DET	0.336	15.9
Kevin Love	CLE	0.400	15.4
Julius Randle	NOP	0.318	14.1
Pascal Siakam	TOR	0.383	13.8
Lauri Markkanen	CHI	0.348	13.1
Kyle Kuzma	LAL	0.363	12.6
Thaddeus Young	IND	0.403	12.2
John Collins	ATL	0.299	12.1
Draymond Green	GSW	0.323	11.7

mentale anche dal punto di vista offensivo, poiché è proprio da quest'ultime che nasce l'opportunità di realizzare un canestro.

2.4.2 Valutazione dei giocatori stratificata per ruolo

Dopo aver valutato i risultati a livello globale, vengono ora presentati quelli stratificati in base al ruolo di appartenenza. Per ciascun ruolo, verranno quindi selezionati i 10 giocatori che hanno ottenuto, con il metodo presentato in questo capitolo, il *ranking* corretto per il rapporto minuti su partite più alto.

Nelle Tabelle 2.6, 2.7, 2.8, 2.9, 2.10 vengono presentati i risultati rispettivamente per i ruoli: ala grande, centro, playmaker, guardia tiratrice e ala piccola.

Dalle tabelle, si può notare che utilizzando questo livello di dettaglio emerge l'importanza di particolari giocatori che nella classifica presentata in precedenza non comparivano, come ad esempio Karl-Anthony Towns, Kevin Durant e Kawhi Leonard, che hanno infatti disputato un'ottima stagione, venendo di fatto inseriti tra i migliori cestisti da molti esperti e da vari siti che si occupano di valutare i giocatori (ESPN, 2019).

Tabella 2.7: Migliori 10 giocatori di ruolo centro classificati dal modello

Giocatore	Squadra	Ranking	Ranking corretto
Anthony Davis	NOP	0.416	22.1
Nikola Jokic	DEN	0.398	20.2
Joel Embiid	PHI	0.311	18.5
Karl-Anthony Towns	MIN	0.329	17.1
Jusuf Nurkic	POR	0.386	16.3
DeMarcus Cousins	GSW	0.396	15.7
Nikola Vucevic	ORL	0.326	14.4
Hassan Whiteside	MIA	0.393	13.7
Andre Drummond	DET	0.277	13.2
Serge Ibaka	TOR	0.401	12.5

Tabella 2.8: Migliori 10 giocatori di ruolo play maker classificati dal modello

Giocatore	Squadra	Ranking	Ranking corretto
James Harden	HOU	0.370	23.0
Russell Westbrook	OKC	0.322	19.6
Trae Young	ATL	0.404	17.4
Stephen Curry	GSW	0.386	17.1
Kemba Walker	CHA	0.355	15.8
Damian Lillard	POR	0.340	15.3
Mike Conley	MEM	0.379	15.2
Ben Simmons	PHI	0.343	15.2
De'Aaron Fox	SAC	0.361	14.3
Lonzo Ball	LAL	0.464	14.1

Tabella 2.9: Migliori 10 giocatori di ruolo guardia tiratrice classificati dal modello

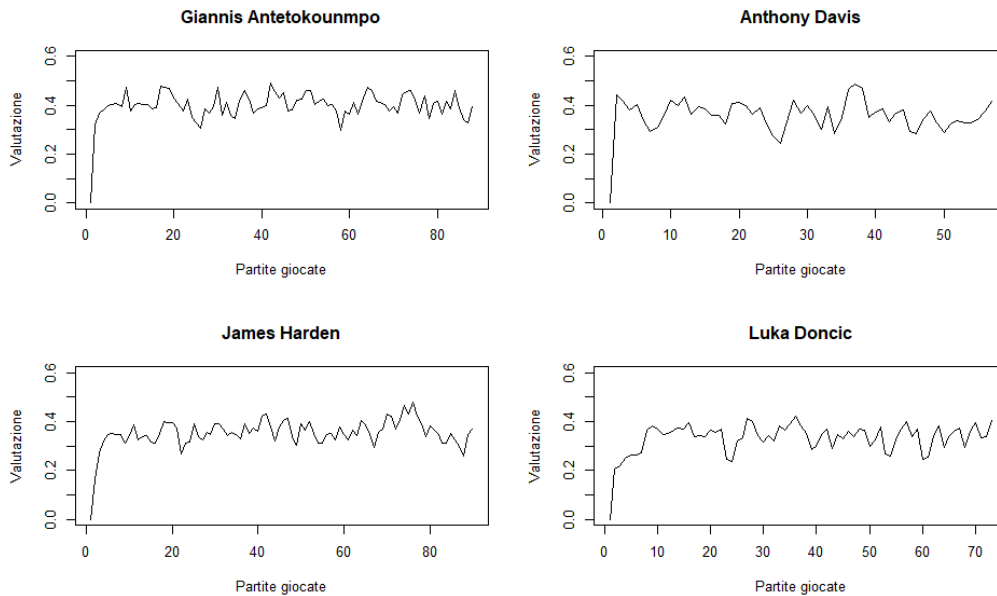
Giocatore	Squadra	Ranking	Ranking corretto
Luka Doncic	DAL	0.407	19.1
Jrue Holiday	NOP	0.395	17.6
Buddy Hield	SAC	0.489	16.7
Bradley Beal	WAS	0.365	16.4
DeMar DeRozan	SAS	0.365	16.0
Devin Booker	PHX	0.301	14.6
Donovan Mitchell	UTA	0.312	13.7
Zach LaVine	CHI	0.324	13.7
Klay Thompson	GSW	0.362	12.6
CJ McCollum	POR	0.333	12.2

Tabella 2.10: Migliori 10 giocatori di ruolo ala piccola classificati dal modello

Giocatore	Squadra	Ranking	Ranking corretto
LeBron James	LAL	0.353	19.1
Paul George	OKC	0.342	17.4
Kawhi Leonard	TOR	0.352	16.5
Kevin Durant	GSW	0.359	16.3
Jimmy Butler	PHI	0.401	13.2
Justise Winslow	MIA	0.409	13.1
Brandon Ingram	LAL	0.353	12.9
Danilo Gallinari	LAC	0.360	12.7
Andrew Wiggins	MIN	0.344	12.1
Khris Middleton	MIL	0.321	12.0

A questo punto, come ultima analisi prima di concludere il capitolo, sembra interessante valutare graficamente l'andamento delle performance dei migliori giocatori per ruolo identificati in precedenza nel corso di tutta la stagione considerata. Nel seguito non saranno presentate le analisi per LeBron James in quanto già presenti in Figura 2.2.

Figura 2.4: Andamento delle performance dei migliori giocatori nel corso della stagione



Dai grafici riportati in Figura 2.4 si può notare come Giannis Antetokounmpo e James Harden siano stati costanti per tutta la durata di stagione, facendo oscillare poco le valutazioni sul proprio conto, con il primo che mostra un rendimento più alto rispetto al secondo per quasi tutto il campionato, eccezion fatta per quelle partite dove James Harden ha segnato un numero molto elevato di punti. Dai grafici emergono poi anche le ottime prestazioni di Anthony Davis e Luka Doncic, i quali però hanno mostrato un andamento più altalenante rispetto ai primi due giocatori.

2.4.3 Valutazione delle performance di squadra

Dopo aver fornito delle valutazioni sui vari giocatori, è stato quindi deciso di proseguire le analisi con lo scopo di utilizzare il *ranking* appena calcolato per definire una classifica delle squadre più pericolose dal punto di vista offensivo.

Per farlo, sono stati utilizzati i dati a disposizione dall'inizio della *Regular Season* fino alla fine dei *playoff*. In particolare, è stato deciso di analizzare quanto una squadra sia stata pericolosa nell'arco di tutta la stagione facendo una media delle valutazioni finali ottenute sui giocatori. Indicando quindi con $\widehat{\text{FinalRank}}_{i,s}$ la stima della valutazione delle performance del giocatore i appartenente alla squadra s ottenuta dopo l'ultima giornata dei *playoff*, si può dunque procedere al calcolo delle performance per ogni squadra s , con $s = 1, \dots, 30$, usando la formula

$$\widehat{\text{TeamRank}}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} \widehat{\text{FinalRank}}_{i,s} \quad (2.4)$$

dove n_s rappresenta il numero totale di giocatori appartenenti alla squadra s .

La formula (2.4) è stata applicata basandosi sulle valutazioni dei 200 giocatori con il *ranking* individuale più alto, portando ai risultati mostrati in Tabella 2.11. Il calcolo, è stato effettuato utilizzando un numero più basso di giocatori per evitare che cestisti con valutazioni particolarmente basse e che hanno disputato solamente gare marginali influenzino eccessivamente i risultati. È comunque interessante notare che la formula precedente può essere calcolata in qualsiasi giornata di campionato, consentendo quindi di fornire anche delle valutazioni dinamiche per le singole squadre.

Dai risultati presenti in Tabella 2.11 si può notare che la squadra che la metodologia proposta classifica come più pericolosa, è quella degli *Oklahoma Thunder*, squadra con la percentuale di palle rubate per gara più elevata. Tuttavia, sebbene il fatto di rubare molti palloni sia fondamentale nel basket, il primo posto non sembra particolarmente sensato, poiché altre squadre si sono comportate meglio dal punto di vista offensivo nel corso della stagione. Al secondo posto si trovano invece i *Golden State Warriors*, squadra che ha disputato un'ottima stagione, arrivando seconda alle NBA *finals*. Questa

Tabella 2.11: Classifica delle performance offensive delle squadre NBA

Squadra	<i>Ranking</i>	Squadra	Ranking
OKC	14.00	BOS	10.10
GSW	13.49	ATL	10.03
MIL	12.97	CHA	10.02
HOU	12.85	DAL	10.02
DET	12.66	DEN	9.94
NOP	11.89	MIN	9.93
SAC	11.38	IND	9.80
PHI	11.16	MIA	9.79
TOR	11.15	LAC	9.64
CLE	11.02	SAS	9.23
WAS	10.81	PHX	8.96
CHI	10.74	ORL	8.86
UTA	10.72	MEM	8.55
LAL	10.63	BKN	8.43
POR	10.39	NYK	8.08

squadra ha inoltre ottenuto la percentuale di stoppate e di assist per gara più alta di tutte le altre squadre NBA, venendo di fatto considerata una delle più pericolose a livello offensivo. Al terzo posto si trovano poi i *Milwaukee Bucks*, squadra che con un gran gioco è riuscita ad avere la media punti e il numero di rimbalzi per gara più elevato di tutte le altre squadre.

Nelle posizioni successive si trovano poi altre squadre che sono state protagoniste o durante la *Regular Season*, come i *Los Angeles Lakers*, o nei *playoff*, come i *Toronto Raptors*. I risultati riportati sembrano quindi soddisfacenti da un punto di vista qualitativo dopo il confronto con alcune classifiche stilate utilizzando degli indici riassuntivi delle performance di squadra (Basketball-reference, 2019). Tuttavia, lo scopo della presente tesi è principalmente quello di valutare le performance dei singoli giocatori, e di conseguenza la valutazione delle squadre potrebbe essere approfondita maggiormente in futuro.

2.5 Valutazioni finali

In conclusione l'applicazione di un modello di regressione logistica stimato con il metodo di riduzione della distorsione di Firth comprendente tutte le variabili che descrivono l'azione e gli effetti fissi relativi al ruolo e ai singoli giocatori, e successivamente l'utilizzo della statistica EWMA per aggregare i risultati delle varie partite, hanno portato alla costruzione di una valutazione delle performance offensive dei giocatori che sembra essere soddisfacente. Confrontando infatti i risultati con gli indicatori disponibili sui siti precedentemente citati si osservano per lo più conclusioni simili, facendo emergere come questo approccio basato sui dati e che utilizza tecniche statistiche basate sulle azioni all'interno di una partita permetta di fornire delle valutazioni attendibili.

Superato anche il problema di natura computazionale, uno dei problemi del metodo presentato riguarda la capacità che il modello utilizzato ha di classificare le azioni in relazione alla variabile risposta. Seppure infatti questa sia una *proxy* del vero oggetto d'interesse, sarebbe preferibile utilizzare un modello che consenta una classificazione migliore dei giocatori e che riduca il numero di falsi positivi e di falsi negativi.

Al di là di questo aspetto che potrà essere oggetto di studi futuri, i risultati ottenuti con questo tipo di approccio sembrano fornire dei buoni risultati, consentendo di effettuare varie valutazioni sui giocatori e di offrire delle indicazioni alle squadre sia nel corso della stagione che durante la sessione di mercato.

Capitolo 3

Stima *bayesiana* del modello logistico per la valutazione dei giocatori NBA

Dopo aver affrontato il problema utilizzando un modello di regressione logistica stimato con il metodo di riduzione della distorsione di Firth, viene ora presentato un approccio alternativo per valutare le performance offensive dei giocatori di NBA. Questo prevede sempre di utilizzare delle tecniche di modellazione per prevedere le probabilità che ha ciascuna azione di essere considerata pericolosa, ma questa volta mediante un modello *bayesiano*.

L'aspetto principale che contraddistingue a livello concettuale i metodi *bayesiani*, i quali hanno avuto una notevole diffusione negli ultimi anni congiuntamente all'aumento delle capacità di calcolo dei computer, sta nel fatto che in questi metodi non si crede che esista un vero valore del parametro oggetto di studio ma si presuppone che i parametri seguano una distribuzione di probabilità, non venendo considerati come delle costanti ma bensì come delle quantità aleatorie.

Questi approcci, che verranno presentati in modo dettagliato nel seguito, hanno le loro fondamenta nel teorema di Bayes (si veda ad esempio Efron, 2013), il quale prevede che, data una partizione dello spazio degli eventi B_1, \dots, B_J tali che $P(B_i) \neq 0$ per $i = 1, \dots, J$, e dato un evento A , la

probabilità di B_j , condizionata ad A , può essere calcolata come

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^J P(A|B_j)P(B_j)}.$$

Questo teorema è utilizzato per aggiornare l'informazione a priori in informazione a posteriori, processo che sta alla base dell'inferenza *bayesiana*.

3.1 L'inferenza *bayesiana*

In questa sezione del capitolo verranno quindi presentati formalmente i vari aspetti dell'inferenza *bayesiana*. Lo scopo principale di questa tipologia di approccio è quello di stimare una distribuzione detta *a posteriori* per il parametro del modello, condizionatamente ai dati osservati. Per fare ciò, è però necessario specificare una distribuzione detta *a priori* per il parametro, la quale dovrà sintetizzare l'informazione che si ha a disposizione riguardo il parametro stesso prima dell'acquisizione dei dati da analizzare.

Formalmente, avendo a disposizione un'insieme di osservazioni $\mathbf{y} = (y_1, \dots, y_n)$, si assume che queste siano realizzazioni indipendenti delle variabili casuali Y_1, \dots, Y_n , ciascuna avente funzione di densità $f_Y(y_i|\theta)$ per $i = 1, \dots, n$, dove θ rappresenta il parametro oggetto di studio. Successivamente, utilizzando congiuntamente l'informazione contenuta nei dati e quella codificata in forma probabilistica su θ , quello che si vuole fare è trovare la distribuzione $\pi(\theta|\mathbf{y})$. Più in dettaglio, si può definire la funzione di verosimiglianza come

$$L(\theta; \mathbf{y}) = f_Y(y_1, y_2, \dots, y_n|\theta) = \prod_{i=1}^n f_{Y_i}(y_i|\theta),$$

dove l'ultima uguaglianza vale se le componenti di \mathbf{y} sono indipendenti ed identicamente distribuite.

Dopodiché, una volta scelta la distribuzione a priori per i parametri d'interesse, indicata con $\pi(\theta)$, si può ricavare la distribuzione a posteriori sfruttando il teorema di Bayes precedentemente menzionato. Questo viene fatto utilizzando la formula

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\theta)L(\theta; \mathbf{y})}{f(\mathbf{y})} = \frac{\pi(\theta)L(\theta; \mathbf{y})}{\int_{\Theta} \pi(\theta)L(\theta; \mathbf{y})d\theta}. \quad (3.1)$$

In particolare, quello che viene fatto con questo teorema corrisponde ad un aggiornamento della distribuzione a priori sfruttando l'informazione contenuta nei dati.

Nel dettaglio, quando si decide di utilizzare un approccio di questo tipo, quello che bisogna fare inizialmente, è scegliere la distribuzione a priori $\pi(\theta)$. Questa deve sintetizzare l'informazione che si ha sul parametro prima di aver eseguito l'esperimento, sfruttando quindi informazioni note sulla base di conoscenze precedenti.

La scelta di questa distribuzione è particolarmente rilevante, e può essere fatta seguendo diverse strategie. La prima è quella di utilizzare delle conoscenze pregresse sul fenomeno oggetto di studio. La seconda è quella di utilizzare come distribuzione a priori una distribuzione *coniugata*, ovvero scegliere una famiglia per $\pi(\theta)$ tale che $\pi(\theta|\mathbf{y})$ appartenga alla stessa famiglia, ottenendo così molti vantaggi dal punto di vista computazionale. Il terzo modo per scegliere la distribuzione a priori è quello di utilizzare delle distribuzioni non informative, come ad esempio la distribuzione a priori di Jeffreys o altre distribuzioni con una varianza particolarmente elevata, che consentono di utilizzare i metodi legati all'inferenza bayesiana senza dover effettuare delle assunzioni particolarmente stringenti sui parametri. La specificazione di questa distribuzione riveste quindi un ruolo importante, poiché per diverse scelte di $\pi(\theta)$ si otterranno diversi risultati, portando eventualmente anche a diverse conclusioni.

Una volta scelta la distribuzione a priori si può quindi procedere al calcolo della posteriori, la quale rappresenta l'obiettivo dell'inferenza e contiene al suo interno tutta l'informazione d'interesse. Questa permette ad esempio di ottenere delle stime dei parametri, di ricavare degli intervalli di probabilità, di effettuare dei test o di calcolare delle previsioni. In particolare, per ottenere le stime dei parametri necessarie per affrontare il problema e poi trarre delle conclusioni, si procede andando ad utilizzare un qualsiasi indice che permetta di riassumere l'informazione contenuta nella distribuzione a posteriori, come ad esempio la moda o il valore atteso.

Tuttavia, ci sono alcuni aspetti che complicano e limitano particolarmente l'utilizzo di queste tecniche inferenziali. Quello più rilevante è rappresentato dai problemi di natura computazionale che sorgono sia nel calcolo della a

posteriori, che anche nella stima dei parametri. Per ottenere $\pi(\theta|\mathbf{y})$ infatti, è necessario calcolare l'integrale presente al denominatore del membro di destra della formula (3.1), il che può risultare decisamente complicato quando la dimensione dello spazio parametrico è elevata. Inoltre, a parte la moda a posteriori, tutti gli altri indici riassuntivi che possono essere utilizzati per riassumere l'informazione a posteriori richiedono di risolvere degli integrali, i quali si complicano quando la dimensionalità del problema è alta.

Per risolvere i problemi computazionali legati all'integrazione numerica, si possono percorrere principalmente due strade. La prima è quella di utilizzare come priori le distribuzioni coniugate alla verosimiglianza ottenuta nel problema. La seconda è quella di utilizzare dei metodi di simulazione per approssimare la distribuzione a posteriori. Per quanto riguarda la prima metodologia, questa non può sempre essere applicata, poiché talvolta la distribuzione a priori coniugata potrebbe non esistere, o anche nel caso in cui questa esistesse, potrebbe non essere adatta per affrontare il problema. La seconda metodologia invece, prevede di utilizzare delle tecniche Monte Carlo, simulando quindi un elevato numero di valori dalla posteriori e stimando le quantità teoriche di interesse con le corrispondenti quantità empiriche ottenute dai valori simulati. Ad esempio, se la quantità d'interesse fosse il valore atteso di una qualche funzione dei parametri, calcolabile utilizzando la distribuzione a posteriori come

$$E[f(\theta)] = \int_{\Theta} f(\theta)\pi(\theta|y)d\theta,$$

questa potrebbe essere approssimata tramite l'utilizzo di tecniche Monte Carlo sfruttando una serie di B valori simulati indipendenti ed identicamente distribuiti di θ provenienti da $\pi(\theta|y)$ utilizzando la formula

$$\widehat{E[f(\theta)]} = \frac{1}{B} \sum_{i=1}^B f(\theta_i).$$

Per procedere in questo modo, si ha tuttavia la necessità di dover simulare valori dalla distribuzione a posteriori, cosa che può essere complicata, in particolare quando la dimensionalità del parametro d'interesse è elevata.

3.1.1 *Markov Chain Monte Carlo*

Nel seguito verrà presentato uno dei metodi più utilizzati per simulare dalla distribuzione a posteriori. In generale, per simulare valori da una distribuzione, una buona tecnica potrebbe essere quella di utilizzare l'algoritmo di accettazione-rifiuto oppure l'*importance sampling*; queste tecniche tuttavia non sono particolarmente semplici da implementare quando, come nel caso della presente tesi, la dimensionalità del problema è particolarmente elevata. Per questo motivo, si ha quindi la necessità di utilizzare un altro metodo, noto come *Markov Chain Monte Carlo* (MCMC, per il quale si veda ad esempio Gamerman e Lopes, 2006). Questo consente di simulare dei valori da una distribuzione dalla quale non risulterebbe semplice simulare utilizzando altri metodi, sfruttando le proprietà delle catene di Markov.

Tuttavia, bisogna tenere in considerazione che i valori simulati da una catena di Markov sono per loro natura dipendenti, e non corrispondono quindi ad un campione casuale semplice dalla distribuzione obiettivo. Ciò nonostante, si ha che se vengono simulati un numero N di valori abbastanza elevato, dopo un periodo di riscaldamento noto come *burnin* (o *warmup*), questi possono essere utilizzati per studiare le proprietà della distribuzione.

Gli algoritmi più noti per fare MCMC sono il Metropolis-Hastings (per il quale si veda ad esempio Chib e Greenberg, 1995) e il *Gibbs Sampling* (per il quale si veda ad esempio Gamerman e Lopes, 2006, Paragrafo 5). Tuttavia, non è sempre facile trovare delle implementazioni efficienti di tali algoritmi in modelli complessi. Una recente alternativa, chiamata *Hamiltonian Monte Carlo*, si è recentemente imposta all'attenzione grazie anche al software STAN che ne dà un'implementazione generale ed efficiente. Questo metodo verrà brevemente introdotto nel paragrafo seguente.

3.2 *Hamiltonian Monte Carlo*

Nel seguito verrà presentato brevemente il metodo noto come *Hamiltonian Monte Carlo* (HMC) (Duane *et al.*, 1987). Verranno prima elencate le motivazioni che hanno portato a dover utilizzare questo particolare algoritmo e successivamente verrà presentato il metodo nelle sue generalità.

3.2.1 Limiti dei metodi precedenti

I metodi MCMC classici, quali il Metropolis-Hastings o il *Gibbs Sampling* presentano alcuni limiti, che sebbene in ambito uni-dimensionale o nei casi in cui la dimensionalità del problema non è elevata non sono così rilevanti, quando la dimensione o la complessità dei modelli utilizzati aumenta si aggravano particolarmente.

Per quanto riguarda l'algoritmo di Metropolis e Hastings, si ha che questo denota alcuni limiti soprattutto per quanto riguarda gli aspetti computazionali. L'utilizzo di questo metodo può infatti portare a dover effettuare un gran numero di iterazioni prima di raggiungere la convergenza. Inoltre, bisogna considerare che anche i più efficienti metodi di questa classe portano ad accettare i valori che ad ogni iterazione dell'algoritmo vengono proposti con una percentuale di circa il 25% (Gelman *et al.*, 1997).

Per quanto riguarda il *Gibbs Sampling*, si ha invece che questo risulta generalmente più efficiente rispetto all'algoritmo Metropolis Hastings, tuttavia ha il grosso svantaggio di richiedere la conoscenza delle densità condizionate e la possibilità di simulare da esse. Oltre a questo aspetto, si ha inoltre che anche utilizzando questa metodologia, la convergenza potrebbe risultare abbastanza lenta quando si è in un modello che coinvolge un gran numero di parametri, portando a dei problemi di efficienza (si veda ad esempio Robert, 2001).

L'*Hamiltonian Monte Carlo* sembra ovviare i problemi dei metodi precedenti.

3.2.2 Il metodo generale

L'idea di base del metodo è quella di aumentare l'efficienza dell'algoritmo Metropolis Hastings sfruttando il gradiente della funzione di densità a posteriori per far sì che i valori simulati delle catene vadano verso regioni con una densità a posteriori più alta. In questo modo, si otterrà quindi il risultato di aumentare il tasso con cui i valori proposti vengono accettati, con un conseguente aumento di efficienza rispetto ai metodi precedenti. Inoltre, usando questa metodologia si ha che sebbene vengano spesso campionati valori con

un'alta densità a posteriori, le code della distribuzione d'interesse vengono comunque rappresentate adeguatamente.

Venendo ora alla spiegazione dettagliata del metodo, si può inizialmente affermare, per fornire un'intuizione riguardo al suo funzionamento, che l'HMC considera il problema di campionamento come l'esplorazione di un sistema fisico, con $-\log[\pi(\theta|\mathbf{y})]$ che rappresenta l'energia potenziale in un punto θ . Più precisamente essendo che per campionare i valori viene utilizzato il gradiente della funzione di densità a posteriori, i movimenti che vengono effettuati dall'algoritmo per raggiungere i punti con una densità più elevata, vengono descritti dalle equazioni Hamiltoniane, le quali sono ampiamente utilizzate in meccanica per descrivere gli scambi di energia cinetica e potenziale che determinano la posizione di un oggetto in un dato momento (per ulteriori dettagli si veda ad esempio Neal, 2011, Paragrafo 5).

Formalmente, indicando con θ un vettore che segue la distribuzione d'interesse $\pi(\theta|\mathbf{y})$, e con $\boldsymbol{\gamma}$ un parametro ausiliario che verrà poi utilizzato nelle equazioni Hamiltoniane per simulare valori da $\pi(\theta|\mathbf{y})$, si può scrivere la funzione Hamiltoniana come: $H(\theta, \boldsymbol{\gamma}) = W(\theta) + K(\boldsymbol{\gamma})$, con $\theta, \boldsymbol{\gamma} \in \mathbb{R}^k$. Questa funzione corrisponde, seguendo l'intuizione precedente, alla somma tra energia potenziale ed energia cinetica. A questo punto, definendo $W(\theta) = -\log[\pi(\theta|\mathbf{y})]$ e ipotizzando per $\boldsymbol{\gamma}$ una distribuzione normale multivariata tale che $\boldsymbol{\gamma} \sim N_k(\mathbf{0}, M)$, con M matrice di covarianza nota, si può riscrivere $H(\theta, \boldsymbol{\gamma})$ come

$$H(\theta, \boldsymbol{\gamma}) = -\log[\pi(\theta|\mathbf{y})] + \frac{1}{2}\boldsymbol{\gamma}^T M^{-1}\boldsymbol{\gamma}.$$

Una volta definita questa equazione, per simulare valori θ provenienti da $\pi(\theta|\mathbf{y})$, si può procedere simulando valori θ e $\boldsymbol{\gamma}$ tramite il tempo fittizio τ usando le equazioni Hamiltoniane derivanti da $H(\theta, \boldsymbol{\gamma})$

$$\begin{cases} d\theta = M^{-1}\boldsymbol{\gamma}d\tau, \\ d\boldsymbol{\gamma} = \nabla_{\theta} \log[\pi(\theta|\mathbf{y})]d\tau \end{cases} \quad (3.2)$$

e scartando successivamente i valori $\boldsymbol{\gamma}$.

Le equazioni (3.2) sono tuttavia particolarmente complesse da risolvere e di conseguenza, viene utilizzato un metodo numerico noto come *leapfrog* (per approfondire si veda ad esempio Girolami e Calderhead, 2011) che fornisce

risultati approssimati in un numero di *step* pari a L . Una volta definito un valore ϵ , in ciascun passo dell'algoritmo vengono quindi iterate le seguenti operazioni per simulare valori di θ :

$$\begin{aligned}\gamma_{\tau+\frac{\epsilon}{2}} &= \gamma_{\tau} + \frac{\epsilon}{2} \nabla_{\theta_{\tau}} \log[\pi(\theta_{\tau}|\mathbf{y})], \\ \theta_{\tau+\epsilon} &= \theta_{\tau} + \epsilon M^{-1} \gamma_{\tau+\frac{\epsilon}{2}}, \\ \gamma_{\tau+\epsilon} &= \gamma_{\tau+\frac{\epsilon}{2}} + \frac{\epsilon}{2} \nabla_{\theta_{\tau+\epsilon}} \log[\pi(\theta_{\tau+\epsilon}|\mathbf{y})].\end{aligned}$$

Dopodiché, essendo che questo metodo fornisce dei risultati approssimati, viene introdotto un passaggio di accettazione-rifiuto, che ha l'obiettivo di garantire che la funzione nei valori simulati proposti non si discosti troppo da $H(\theta, \gamma)$.

Nelle equazioni definite in precedenza, i valori ϵ e L utilizzati sono dei parametri di regolazione, i quali devono essere scelti in maniera adeguata affinché l'algoritmo possa fornire buoni risultati. Generalmente, si ha che scegliendo valori di ϵ piccoli, si avranno approssimazioni delle soluzioni delle equazioni Hamiltoniane più vicine a quelle reali e quindi tassi di accettazione dei valori proposti più elevati. Tuttavia, scegliendo ϵ piccolo, bisogna scegliere un valore di L abbastanza grande, in modo tale che ϵL sia sufficientemente elevato da consentire di generare valori anche sulle code della distribuzione. Come ultimo aspetto per quanto riguarda i parametri di regolazione, si ha che ϵL non deve comunque essere troppo grande, poiché altrimenti l'algoritmo perderebbe di efficienza. Un buon metodo per selezionare questi parametri (si veda ad esempio Neal, 2011) è fare in modo che il tasso di accettazione dei valori proposti sia pari al 65% circa.

Definiti anche questi ultimi aspetti, viene quindi presentato l'Algoritmo 1, che implementa l'*Hamiltonian Monte Carlo* per generare valori da $\pi(\theta|\mathbf{y})$.

In dettaglio, nell'Algoritmo 1, con *Leapfrog* viene indicato il metodo numerico con cui vengono risolte le equazioni Hamiltoniane, il quale permette di effettuare un aggiornamento dei valori di θ e γ proposti.

Tramite questo metodo, si può dunque procedere utilizzando una procedura *bayesiana* per affrontare problemi di natura complessa anche nel caso di dimensionalità elevate, facendo sì che le catene di Markov simulate per stimare poi vari aspetti dei parametri, siano in grado di raggiungere

Algoritmo 1: *Hamiltonian Monte Carlo***Obiettivo:** Simulare N valori da una distribuzione $\pi(\cdot|\mathbf{y})$;Valori in ingresso: $\theta^{(0)}$, M , ϵ, L ; $N = 10^3$;**for** $t = 1, \dots, N$ **do** $\gamma = N(0, M)$; $\theta^{(t)} = \theta^{(t-1)}$, $\tilde{\theta} = \theta^{(t-1)}$, $\tilde{\gamma} = \gamma$; **for** $l = 1, \dots, L$ **do** $\tilde{\theta}, \tilde{\gamma} \leftarrow \text{Leapfrog}(\tilde{\theta}, \tilde{\gamma}, \epsilon, M)$ **end** $\alpha = \min\left(1, \frac{\exp(\log[\pi(\tilde{\theta}|\mathbf{y})] - \frac{1}{2}\tilde{\gamma}^T M \tilde{\gamma})}{\exp(\log[\pi(\theta^{(t-1)}|\mathbf{y})] - \frac{1}{2}\gamma^T M \gamma)}\right)$; Con probabilità α : $\theta^{(t)} = \tilde{\theta}$;**end****Ritorna:** $\theta^{(1)}, \dots, \theta^{(N)}$;

la convergenza più velocemente rispetto ad altri algoritmi, come ad esempio il Metropolis-Hastings e il *Gibbs Sampling*, fornendo generalmente anche risultati più affidabili.

3.3 Inferenza bayesiana nel modello logistico

Dopo aver presentato l'HMC, viene ora descritta una sua applicazione al modello logistico utilizzato in questa tesi. Di conseguenza, verrà inizialmente presentata la scelta delle distribuzioni a priori utilizzate, successivamente verrà discussa la costruzione della funzione di densità a posteriori, e infine verrà applicato l'HMC per simulare dei valori da quest'ultima.

3.3.1 Scelta delle distribuzioni a priori e calcolo della distribuzione a posteriori

Formalmente, utilizzando il modello specificato nel paragrafo 2.1, l'obiettivo dell'inferenza è ora quello di stimare le distribuzioni dei parametri conte-

nuti in β associati alle variabili esplicative. Le densità a posteriori ottenute verranno poi utilizzate per prevedere l'ignoto parametro oggetto di studio π_i .

Inizialmente, quello che viene fatto è definire le distribuzioni a priori che verranno utilizzate nelle analisi. Per spiegare il procedimento adottato quando la variabile risposta è dicotomica e successivamente anche nel caso della presente tesi, è stato scelto di utilizzare per ciascun parametro contenuto nel vettore β delle distribuzioni a priori Gaussiane. Ciò corrisponde alla scelta di una distribuzione Gaussiana multivariata per il vettore β tale che $\beta \sim N_p(\mathbf{0}, \sigma_\beta^2 I)$, dove con σ_β^2 viene indicato l'iperparametro associato alla distribuzione a priori, la cui scelta verrà approfondita nel seguito.

Successivamente, si può procedere nelle analisi andando a calcolare la funzione di log-verosimiglianza associata ai dati, la quale può essere calcolata, dopo una serie di passaggi, come

$$l(\beta; X, \mathbf{y}) = \log[f(y|X, \beta)] = \beta^T X^T (\mathbf{y} - \mathbf{1}_n) - \mathbf{1}_n^T [\log(1 + \exp(-x_i \beta))]_{n \times 1}.$$

Definite la funzione di log-verosimiglianza e la densità a priori, si può quindi procedere al calcolo della log-posteriori, che rappresenta il logaritmo della funzione di densità a posteriori. Quindi, ricorrendo al teorema di Bayes, si ottiene che la log-posteriori, a meno della log-costante di normalizzazione, è data da:

$$\log[\pi(\beta|y, X, \sigma_\beta^2)] = \beta^T X^T (\mathbf{y} - \mathbf{1}_n) - \mathbf{1}_n^T [\log(1 + \exp(-x_i \beta))]_{n \times 1} - \frac{\beta^T \beta}{2\sigma_\beta^2}.$$

3.3.2 Stima del modello tramite *HMC*

Per utilizzare l'algoritmo HMC, bisogna prima di tutto definire la densità del parametro γ , per la quale è stato scelto di utilizzare una Gaussiana tale che $\gamma \sim N_k(\mathbf{0}, M)$, con M che, vista l'assenza di informazioni precise sulla distribuzione obiettivo, è stata posta pari alla matrice identità. Successivamente, si può quindi scrivere la funzione Hamiltoniana come

$$H(\beta, \gamma) \propto \log[\pi(\beta|y, X, \sigma_\beta^2)] + \frac{1}{2} \gamma^T M^{-1} \gamma,$$

dalla quale si possono ricavare le equazioni Hamiltoniane precedentemente definite, le quali verranno poi utilizzate per simulare i valori desiderati.

A questo punto, per proseguire con l'implementazione dell'algoritmo, viene definito il gradiente della funzione di densità a posteriori, il quale permetterà poi di utilizzare il metodo *leapfrog* per aggiornare i valori proposti dall'HMC. Il gradiente è pari a

$$\nabla_{\beta} \log[\pi(\beta|y, X, \sigma_{\beta}^2)] = X^T \left[\mathbf{y} - \mathbf{1}_n + \left(\frac{\exp(-\mathbf{x}_i \beta)}{1 + \exp(-\mathbf{x}_i \beta)} \right)_{n \times 1} \right] - \beta / \sigma_{\beta}^2.$$

A questo punto si hanno a disposizione tutte le quantità necessarie per implementare l'Algoritmo 1, ad eccezione dei parametri di regolazione ϵ ed L . Un buon metodo per ottenere quest'ultimi, è quello di utilizzare il metodo NUTS (*no-U-turn sampling*, per approfondimenti si veda ad esempio Hoffman e Gelman, 2014) con il quale questi due parametri vengono regolati in modo adattivo, selezionando L in modo tale da rendere l'algoritmo il più efficiente possibile e successivamente trovando il valore ottimale per ϵ .

Infine, si può quindi procedere con l'inferenza utilizzando i valori simulati tramite l'HMC.

3.3.3 STAN

L'inferenza *bayesiana* in un modello logistico attraverso HMC si può effettuare usando il pacchetto R Rstanarm (Goodrich *et al.*, 2020), il quale permette di definire delle distribuzioni a priori, di utilizzare l'*Hamiltonian Monte Carlo* per simulare valori dalla posteriori e successivamente, sulla base dei valori simulati, di ottenere diverse statistiche di sintesi legate alla posteriori.

Più dettagliatamente, questo pacchetto rappresenta un'interfaccia nel linguaggio di programmazione R al pacchetto STAN (per approfondimenti si veda ad esempio Carpenter *et al.*, 2017) scritto in C++, il quale permette la codifica di modelli di probabilità in ambito *bayesiano* e l'utilizzo di strumenti di analisi a posteriori per valutare i risultati. In particolare, gli strumenti di analisi a posteriori di STAN permettono di calcolare una serie di statistiche riassuntive utilizzando i valori simulati, come la media a posteriori o gli *standard error*. Infine consentono di calcolare delle statistiche di diagnostica per monitorare la convergenza delle catene simulate (per ulteriori dettagli riguardo la studio della convergenza delle catene si veda ad esempio Gelman e Rubin, 1992).

3.3.4 L'importanza delle distribuzioni a priori

Come già detto più volte nel corso del capitolo, una delle prime cose da fare in un'analisi statistica che utilizza metodi *bayesiani*, è quella di selezionare le distribuzioni a priori per i parametri. In molti problemi reali non si hanno tuttavia particolari informazioni riguardo il fenomeno che si sta studiando, e ciò rende difficile selezionare una distribuzione a priori adeguata per affrontare il problema. Questo aspetto potrebbe quindi portare ad utilizzare delle priori "piatte", le quali assegnano una probabilità più o meno simile a tutti i valori del supporto. Tuttavia, procedere in questo modo non sembra molto sensato, poiché così facendo si corre il rischio di attribuire la stessa probabilità sia a valori del supporto che sono ragionevoli per i parametri considerati che anche a valori che non lo sono.

Di conseguenza, anche quando non si hanno particolari informazioni sul fenomeno oggetto di studio, una distribuzione a priori piatta non corrisponderà mai alla scelta migliore da effettuare, portando quindi a preferire delle distribuzioni a priori debolmente informative, le quali permettono sia di utilizzare delle deboli conoscenze sul fenomeno, che di rendere i calcoli più stabili, portando quindi anche a velocizzare il raggiungimento dei risultati in termini di tempi computazionali (Betancourt, 2017).

Nella presente tesi quindi, almeno inizialmente, è stato deciso di utilizzare delle distribuzioni a priori debolmente informative. In particolare, è stato scelto di utilizzare delle distribuzioni Gaussiane con parametro di posizione il vettore nullo e come valori degli iperparametri di scala che definiscono la variabilità, dei valori piccoli, intorno a 2.5. Questi valori corrispondono a quelli di *default* del pacchetto *rstanarm*, il quale effettua comunque un'altra operazione sui parametri di scala scelti, in modo da fornire delle distribuzioni a priori che siano più vicine ai dati, e quindi al fenomeno osservato. Più precisamente, il pacchetto permette di riscaldare opportunamente i valori scelti per quanto riguarda il parametro di scala, in modo tale da tenere conto dell'ordine di grandezza delle variabili. Questo approccio introduce tuttavia una leggera dipendenza dai dati, poiché l'operazione che modifica leggermente il parametro di scala si basa appunto sui valori assunti dai predittori. L'operazione sembra comunque sensata, e porta ad ottenere dei risultati soddisfacenti

nella valutazione delle performance dei giocatori di basket, e pertanto verrà eseguita nel corso delle successive analisi.

3.4 Costruzione del *ranking*

Adesso che sono stati spiegati dettagliatamente i metodi legati all'inferenza *bayesiana* e le sue applicazioni nel caso logistico, si può quindi procedere alla valutazione delle performance offensive dei giocatori NBA sfruttando queste metodologie, che consentiranno infine la costruzione di una classifica dei giocatori come quella riportata nel Capitolo 2. Il *ranking* verrà calcolato inizialmente utilizzando come variabili esplicative tutte le variabili che descrivono l'azione effettuata da ciascun giocatore. Successivamente, verranno introdotti nel calcolo anche gli effetti fissi relativi a ciascun cestista.

L'obiettivo di queste valutazioni rimane il medesimo di quelle fornite nel capitolo precedente, ed è quello di valutare le performance dei giocatori per ogni partita in cui questi sono scesi in campo, considerando però non solo la partita oggetto di studio, ma anche tutte le partite ad essa precedenti, così da includere nella valutazione anche lo stato di forma dei giocatori.

3.4.1 Primi metodi utilizzati

Per procedere quindi alla costruzione del *ranking*, sono stati valutati vari metodi per implementare le tecniche di modellazione.

Il primo procedimento messo in atto per fornire delle valutazioni sulle performance offensive dei giocatori, essendo che l'obiettivo è anche quello di tenere in considerazione lo stato di forma dei cestisti, è stato quello di provare ad adattare un modello per ogni giornata di campionato nella quale è stata disputata almeno una partita, sfruttando tutti i dati disponibili dall'inizio della stagione fino alla giornata considerata. Successivamente, utilizzando la formula (2.1), è stato quindi calcolato il *ranking* dei giocatori giornata dopo giornata sulla base delle probabilità stimate dai modelli *bayesiani*. Questa metodologia, aveva scarsi risultati dal punto di vista dei tempi computazionali anche nel caso delle tecniche proposte nel Capitolo 2. Utilizzando però le metodologie descritte in questo capitolo, le cose peggiorano ulteriormente,

poiché il carico computazionale che comportano questi metodi è di gran lunga maggiore rispetto a quelli precedenti. Per ogni nuova giornata di campionato disponibile bisogna infatti andare a simulare valori dalla distribuzione a posteriori includendo nella simulazione l'informazione relativa ad un numero di osservazioni sempre più alto, provocando quindi un rallentamento notevole nelle operazioni. Questo aspetto è dovuto al fatto che nell'*Hamiltonian Monte Carlo* applicato al caso della regressione logistica, all'aumentare di n i tempi computazionali aumentano molto.

Per sopperire a questo problema si è quindi scelto di utilizzare la stessa metodologia adottata anche nel Capitolo 2, ovvero è stato deciso di calcolare il *ranking* dei giocatori giornata dopo giornata, aggregando i risultati utilizzando la statistica EWMA. Più precisamente, per ogni giornata nella quale è stata giocata almeno una partita, si è deciso di stimare il modello con i dati relativi alla giornata considerata, utilizzando di volta in volta come distribuzioni a priori quelle di *default* definite nel paragrafo precedente. Per ciascun modello stimato, i valori simulati tramite l'HMC sono stati calcolati dalla simulazione di 4 catene (eseguite in parallelo), in ciascuna delle quali vengono simulati 2000 valori, dei quali vengono considerati solamente gli ultimi 1000 (utilizzando quindi un periodo di *warmup* pari a 1000). Successivamente, mediante i valori ottenuti, le analisi sono proseguite andando a stimare i parametri e le probabilità che ogni azione compiuta nel corso di ciascuna giornata sia pericolosa a livello offensivo. Infine, attraverso le formule (2.2), dove viene utilizzato un valore di $\lambda = 0.5$, e (2.3), basate sulle probabilità appena stimate, sono stati quindi calcolati gli indici per valutare le performance offensive dei giocatori.

Nel dettaglio, per effettuare le previsioni tramite questi metodi, vengono utilizzati i valori simulati ottenuti per ogni parametro. In particolare, avendo a disposizione questi valori ripartiti in 4 catene, si potrà quindi sfruttare un numero pari a $B = 4000$ valori provenienti dalla distribuzione a posteriori di ciascun parametro per effettuare le stime. Di conseguenza, per ciascuna azione, si potrà quindi valutare la probabilità che questa risulti importante dal punto di vista offensivo utilizzando i B valori simulati dei parametri per calcolare B valori del predittore lineare, e tramite questi calcolare infine B probabilità. A questo punto, per ottenere la probabilità finale con cui una

determinata azione può essere considerata pericolosa, si può ad esempio fare una media tra le probabilità stimate. Il calcolo è stato sintetizzato nella formula (3.3).

$$\hat{\pi}_i = \frac{1}{B} \sum_{j=1}^B \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}, \quad (3.3)$$

dove con η_{ij} viene indicato il predittore lineare relativo all'azione i ottenuto dai valori simulati j .

Definito anche come sono state calcolate le probabilità, si può quindi ottenere una serie storica delle valutazioni dei giocatori e costruire quindi un *ranking*, come fatto nel capitolo precedente.

3.4.2 Un vantaggio dei metodi *bayesiani*

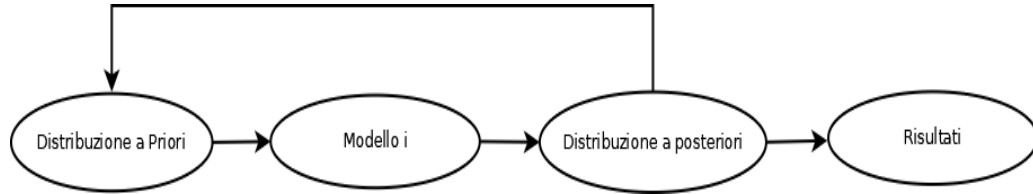
Prima di procedere e mostrare i risultati ottenuti, è tuttavia importante notare che in un approccio *bayesiano* come quello descritto in precedenza si può introdurre una componente di dinamicità ulteriore rispetto a quella introdotta utilizzando solamente la statistica EWMA.

Facendo come prima infatti, ogni modello considera solamente i dati relativi ad una singola giornata, aggregando i risultati solamente dopo aver calcolato le valutazioni relative ad ogni giocatore. Tuttavia, in ambito *bayesiano*, dove per la stima di ogni modello vanno specificate delle distribuzioni a priori, si possono utilizzare i risultati dei modelli stimati nelle giornate precedenti come informazioni d'ingresso per i modelli successivi. Così facendo si avrà che l'informazione riguardante le stime dei parametri di ciascuna partita verrà utilizzata anche nella valutazione delle partite seguenti. Più precisamente, si può quindi effettuare un procedimento che prevede di utilizzare la distribuzione a posteriori dei parametri relativa ad una determinata giornata, come distribuzione a priori per la stima del modello relativo alla giornata successiva, riuscendo così a fornire delle distribuzioni a priori informative, le quali saranno più adeguate rispetto a quelle di *default*.

Il procedimento può essere riassunto dalla Figura 3.1.

Questa metodologia sembra sensata da un punto di vista logico e porta a buoni risultati nel caso della presente tesi, permettendo sia di considerare maggiormente l'andamento a livello prestazionale dei giocatori nel corso della

Figura 3.1: Metodologia adottata per modellare i dati aggregando i risultati delle giornate precedenti



stagione che anche di stabilizzare i calcoli, riducendo i tempi di esecuzione. Tuttavia, è bene prestare un ultimo accorgimento, utilizzando infatti come distribuzione a priori per un modello la distribuzione a posteriori del modello precedente, si rischia di dare troppo peso ad alcune caratteristiche che magari sono emerse in un'unica giornata di campionato, introducendo così una distorsione nei risultati. Per evitare ciò, si può quindi utilizzare come distribuzione a priori di ciascun modello la distribuzione a posteriori risultante dal modello precedente, pesandola però opportunamente con le distribuzioni a priori dei parametri utilizzate in precedenza, così da considerare tutta l'informazione disponibile fino a quel momento. Questo può essere fatto ad esempio utilizzando ancora una volta la statistica EWMA per aggiornare gli iperparametri delle priori. Per quanto riguarda l'iperparametro relativo alla posizione, si può dunque utilizzare la formula

$$\boldsymbol{\mu}_i = (1 - \lambda)\boldsymbol{\mu}_{i-1} + \lambda\widehat{\boldsymbol{\beta}}_{i-1},$$

dove con $\boldsymbol{\mu}_i$ viene indicato il parametro di posizione da utilizzare nella distribuzione a priori per il modello da stimare alla giornata i , mentre con $\widehat{\boldsymbol{\beta}}_{i-1}$ vengono indicate le stime dei coefficienti della distribuzione a posteriori ottenuti nella giornata $i - 1$. Per quanto riguarda l'iperparametro di scala invece, si può utilizzare la stessa formula, sostituendo ai coefficienti stimati gli *standard error* dei parametri ottenuti dal modello.

Questo metodo consente quindi di migliorare i risultati da un punto di vista statistico, poiché porta ad utilizzare come distribuzione a priori delle distribuzioni maggiormente informative, le quali sembrano essere più sensate. Inoltre, consente di fornire delle valutazioni dei giocatori che tengono conto

anche di tutte le partite che questi hanno disputato, rendendo questo metodo preferibile rispetto a quello che utilizzava nei modelli sempre le distribuzioni a priori di *default*.

A questo punto, si può quindi procedere nella valutazione dei giocatori. In particolare, per ciascuna giornata nella quale è stata disputata almeno una partita, dopo aver applicato il modello utilizzando la metodologia descritta, e dopo aver utilizzato quest'ultimo per calcolare le probabilità che ha ciascuna azione di essere considerata pericolosa, vengono utilizzate le formule (2.1), (2.2) e (2.3) per fornire delle valutazioni riguardanti le performance offensive di ciascun giocatore nell'arco di tutta la stagione considerata.

3.4.3 Inserimento dell'effetto fisso

Fino a questo momento, contrariamente a quanto fatto nel capitolo precedente, non è stato considerato nelle analisi l'inserimento di un effetto fisso per ciascun cestista sceso in campo, perdendo quindi l'informazione relativa all'eterogeneità tra i giocatori. Di conseguenza, visti i buoni risultati che gli effetti fissi avevano introdotto nel modello di regressione logistico stimato con il metodo di riduzione della distorsione di Firth, è stato deciso di introdurre anche qui un parametro per ciascun giocatore. Come prima, nella stima di ogni modello e quindi in ogni giornata disponibile, sono stati considerati solamente i giocatori che hanno partecipato attivamente ad almeno 5 azioni. Introducendo gli effetti fissi però, il numero di parametri di cui si vuole trovare la distribuzione a posteriori aumenta di molto, e ciò porta a non raggiungere la convergenza in alcune delle catene simulate, portando a dei campioni di valori simulati la cui dimensione effettiva è troppo bassa per permettere una stima adeguata della distribuzione dei parametri.

Per risolvere questo problema, è tuttavia sufficiente aumentare il numero di iterazioni da far svolgere all'*Hamiltonian Monte Carlo*, così da ottenere delle catene più lunghe. Nel caso in esame è stato quindi deciso di simulare 4 catene, sempre sfruttando la parallelizzazione, ciascuna di 4000 valori, utilizzando un periodo di *burnin* pari a 2000. Questa modifica ha quindi consentito l'inserimento degli effetti fissi dei giocatori, garantendo la convergenza delle catene secondo le statistiche di diagnostica presentate in precedenza.

Tabella 3.1: Tabella riassuntiva delle stime bayesiane dei parametri relativi ai giocatori

Variabile	Media	Standard Error	Intervallo HPD 95%	
			Inferiore	Superiore
Alfonzo McKinnie	0.38	0.54	-0.56	1.26
Andre Iguodala	0.45	0.31	-0.05	0.94
Andrew Bogut	-1.52	0.74	-2.78	-0.35
Danny Green	0.12	0.34	-0.43	0.66
DeMarcus Cousins	-0.76	0.32	-1.29	-0.24
Draymond Green	-0.05	0.24	-0.43	0.33
Fred VanVleet	0.21	0.27	-0.22	0.66
Jonas Jerebko	-0.72	1.74	-3.87	1.95
Kawhi Leonard	0.05	0.18	-0.25	0.34
Kevon Looney	-1.37	0.44	-2.09	-0.66
Klay Thompson	-0.03	0.34	-0.59	0.52
Kyle Lowry	0.17	0.20	-0.15	0.50
Marc Gasol	-0.89	0.23	-1.27	-0.51
Norman Powell	0.30	0.39	-0.35	0.94
Pascal Siakam	0.16	0.21	-0.19	0.52
Quinn Cook	0.26	0.49	-0.57	1.05
Serge Ibaka	-0.87	0.24	-1.27	-0.47
Shaun Livingston	-0.34	0.50	-1.20	0.49
Stephen Curry	0.11	0.25	-0.31	0.52

Il modello di regressione logistica stimato includendo gli effetti fissi dei giocatori e sfruttando l'introduzione di una componente aggiuntiva di dinamicità, correggendo le distribuzioni a priori per le distribuzioni a posteriori ottenute in precedenza, ha portato all'ultima giornata dei *playoff* ad ottenere le stime dei parametri rappresentate nelle Tabelle 3.1 e 3.2. In queste sono riportate oltre alle stime anche gli intervalli di credibilità.

Analizzando nel dettaglio la Tabella 3.2, è particolarmente interessante notare che i tiri che contribuiscono maggiormente a considerare un'azione pericolosa da parte del modello sono le varie categorie dei tiri definite come *Layup*. Il modello porta inoltre a considerare i rimbalzi offensivi e difensivi, così come anche le stoppate, gli assist e la realizzazione di un canestro, come

Tabella 3.2: Tabella riassuntiva delle stime bayesiane dei parametri relativi alle azioni

Variabile	Media	Standard Error	Intervallo HPD 95%	
			Inferiore	Superiore
foul	-1.84	0.03	-1.90	-1.79
freethrow	-0.21	0.01	-0.23	-0.19
jumpball	0.13	0.07	0.00	0.25
shot	-0.43	0.01	-0.44	-0.41
turnover	-0.80	0.02	-0.84	-0.76
violation	-0.06	1.23	-2.20	1.87
AlleyOopDunk	-1.63	1.74	-4.77	1.15
badpass	-1.17	0.07	-1.29	-1.05
DrivingFingerRollLayup	-0.13	0.52	-0.99	0.72
DrivingReverseLayup	1.03	1.38	-1.17	3.43
Dunk	1.55	1.54	-0.83	4.32
FadeawayJumper	-0.16	0.06	-0.26	-0.07
FloatingJumpShot	0.36	0.57	-0.59	1.35
HookShot	0.20	0.60	-0.78	1.15
JumpShot	-0.18	0.01	-0.21	-0.16
Layup	0.17	0.51	-0.69	0.97
lostball	-1.14	0.08	-1.27	-1.00
outofboundslostballturnover	1.63	1.33	-0.66	3.83
PullupJumpShot	-0.18	0.03	-0.22	-0.14
PutbackLayup	-0.67	1.39	-3.06	1.52
rebounddefensive	0.59	0.01	0.57	0.61
reboundoffensive	0.75	0.02	0.73	0.78
ReverseLayup	0.08	1.44	-2.31	2.31
RunningDunk	-0.58	0.75	-1.82	0.65
RunningFingerRollLayup	-0.48	1.39	-2.86	1.71
RunningLayup	0.21	0.25	-0.21	0.63
StepBackJumpShot	-0.27	0.04	-0.33	-0.20
traveling	-1.31	1.72	-4.36	1.24
TurnaroundFadeaway	-0.35	1.05	-2.10	1.38
TurnaroundHookShot	-1.38	1.19	-3.42	0.58
TurnaroundJumpShot	-0.02	0.40	-0.69	0.64
assist	1.69	0.01	1.67	1.71
steal	0.89	0.02	0.85	0.93
block	0.18	0.03	0.13	0.22
ruoloPF	-1.39	0.03	-1.44	-1.33
ruoloPG	-1.32	0.04	-1.38	-1.26
ruoloSF	-1.35	0.04	-1.42	-1.28
ruoloSG	-1.38	0.04	-1.45	-1.32
res	1.94	0.01	1.92	1.95

particolarmente influenti nel rendere importante un'azione a livello offensivo. D'altra parte il modello porta a considerare negativamente commettere dei falli, sbagliare dei passaggi o subire delle ripartenze.

I valori presentati nelle Tabelle 3.1 e 3.2 sono stati ottenuti utilizzando i valori simulati ricavati tramite l'*Hamiltonian Monte Carlo*. In particolare, questo algoritmo ha permesso di simulare dei valori dalla distribuzione a posteriori che portano ad un valore del *Potential Scale Reduction Factor* pari ad 1 per tutti i parametri, valore che indica la raggiunta convergenza nella simulazione delle catene. Più nel dettaglio, l'HMC ha portato ai grafici di diagnostica presentati nelle Figure 3.2, 3.3 e 3.4, che contengono rispettivamente i grafici delle traiettorie, quelli delle autocorrelazioni e i grafici delle distribuzioni a posteriori, i quali permettono di trarre la stessa conclusione delle diagnostiche numeriche. Nelle figure, per questioni di spazio, sono riportati i grafici di diagnostica solamente per i parametri definiti come: *Layup*, *lostball*, *rebound offensive* e *block*. In particolare, nei grafici in Figura 3.2 e 3.3, sono riportati i risultati relativi a ciascuna delle 4 catene simulate.

Arrivati a questo punto, si possono quindi sfruttare i risultati ottenuti dai modelli (stimati in ogni giornata di campionato) per stimare le probabilità che ciascuna azione sia pericolosa mediante la formula (3.3), e successivamente utilizzare le formule (2.2) e (2.3) per calcolare il *ranking* dei giocatori.

3.5 Valutazione delle performance di alcuni giocatori

Adesso che sono stati presentati i metodi per la valutazione delle performance dei cestisti, vengono quindi mostrati i risultati che si sono ottenuti stimando il modello logistico con il metodo *bayesiano*. Come fatto nel capitolo precedente, verranno dunque presentate due tipologie di classifiche dei giocatori NBA: una globale, la quale non effettua alcuna distinzione sulla base del ruolo dei cestisti, e una che invece tiene conto di questa distinzione, permettendo di effettuare dei confronti più ragionevoli e mirati.

3.5 Valutazione delle performance di alcuni giocatori

Figura 3.2: Grafico delle traiettorie dei valori simulati per i parametri identificati come *Layup*, *lostball*, *rebound offensive* e *block* per ciascuna delle 4 catene ottenute

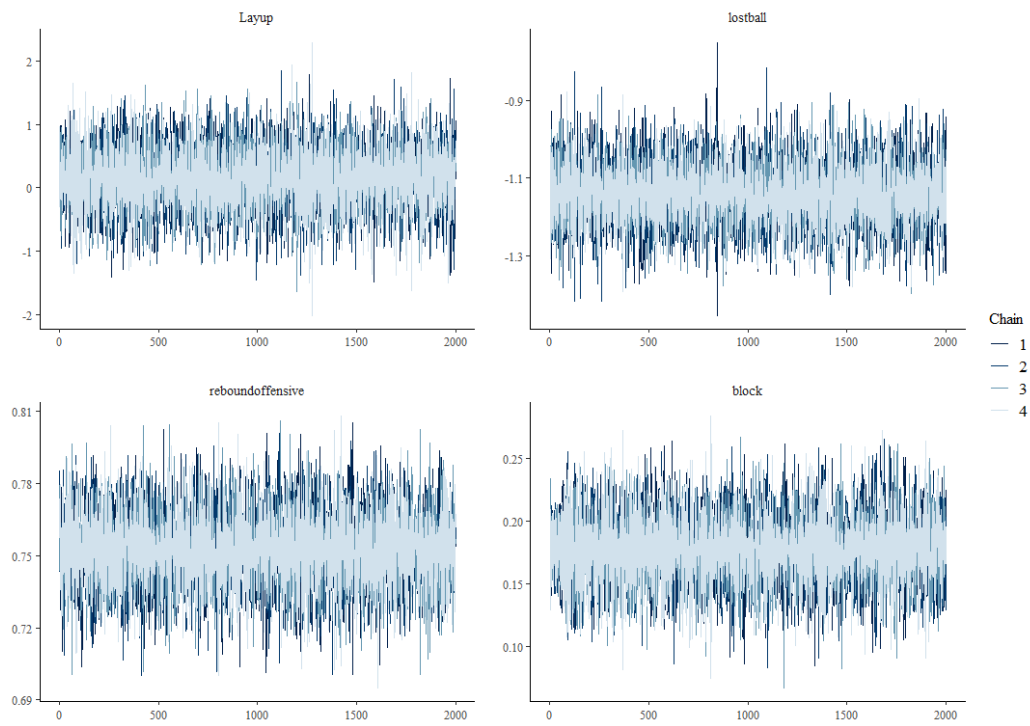
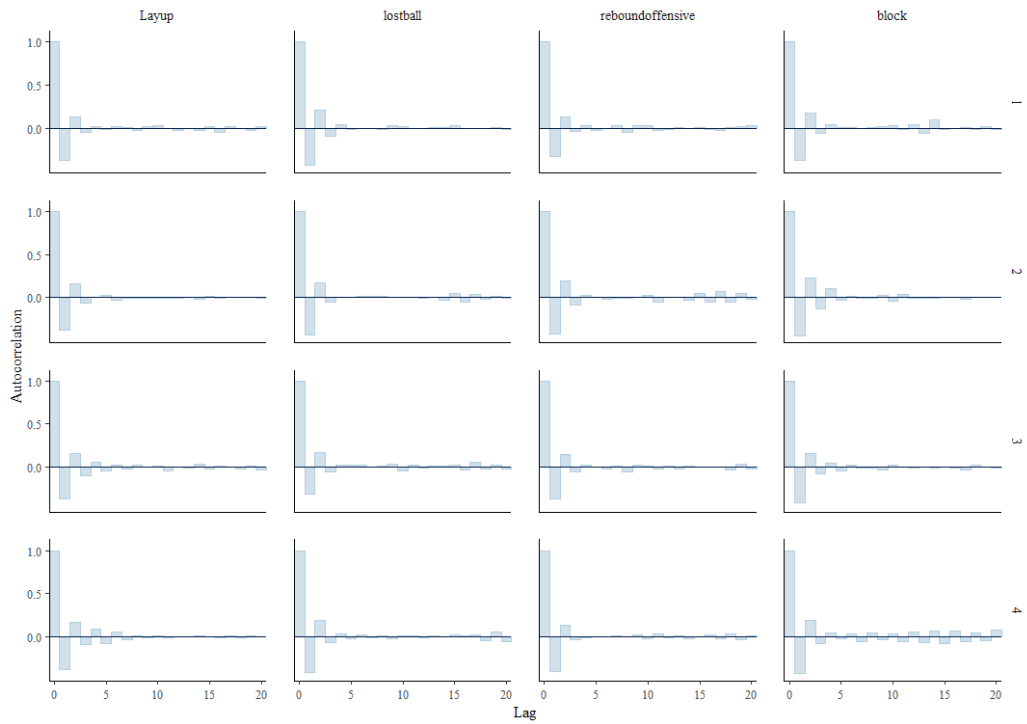


Figura 3.3: Grafico delle autocorrelazioni ottenute dai valori simulati per i parametri identificati come *Layup*, *lostball*, *rebound offensive* e *block* per ciascuna delle 4 catene ottenute



3.5 Valutazione delle performance di alcuni giocatori

Figura 3.4: Grafico delle distribuzioni a posteriori dei parametri identificati come *Layup*, *lostball*, *rebound offensive* e *block*

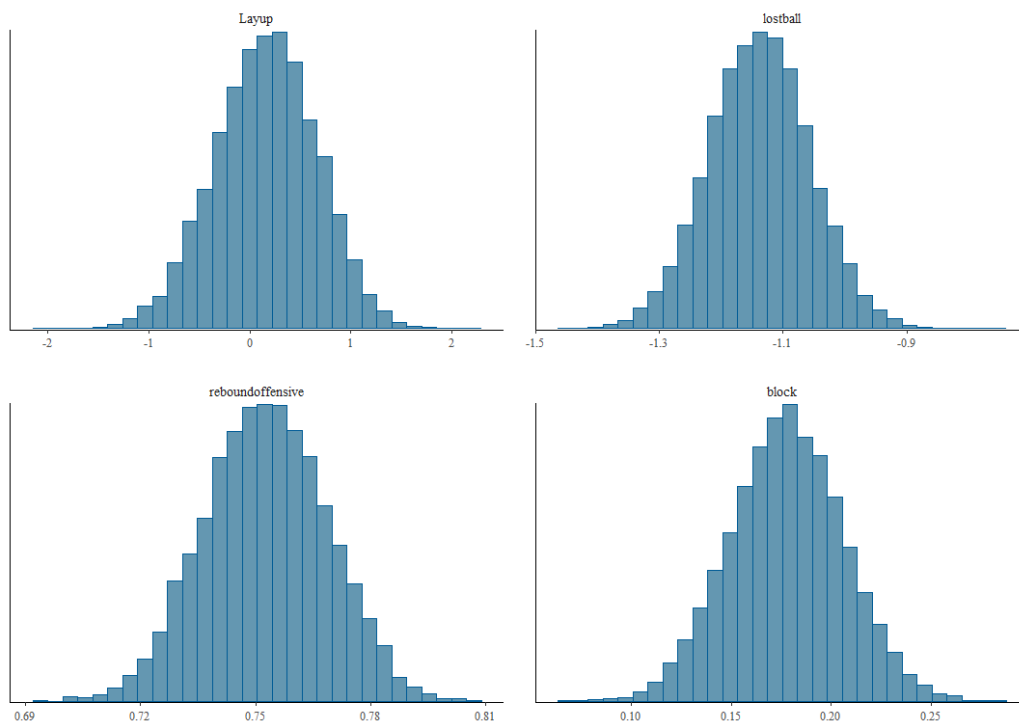


Tabella 3.3: Classifica dei 10 migliori giocatori NBA selezionati dal modello stimato con il metodo bayesiano

Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
Giannis Antetokounmpo	MIL	PF	0.39	23.44
James Harden	HOU	PG	0.36	22.57
Anthony Davis	NOP	C	0.41	21.84
Nikola Jokic	DEN	C	0.39	19.94
Russell Westbrook	OKC	PG	0.32	19.31
Luka Doncic	DAL	SG	0.41	19.03
LeBron James	LAL	SF	0.35	18.83
Joel Embiid	PHI	C	0.30	18.03
Jrue Holiday	NOP	SG	0.39	17.46
Trae Young	ATL	PG	0.40	17.27

Per la costruzione di queste classifiche verrà utilizzato il metodo che utilizza come distribuzioni a priori per una data giornata le posteriori ottenute nella giornata precedente, permettendo così di introdurre un elemento di dinamicità ulteriore rispetto al metodo che utilizza sempre le distribuzioni a priori di *default*.

Infine, verrà presentata anche una valutazione delle performance di squadra. In particolare, utilizzando i risultati ottenuti sui singoli giocatori, si andrà a fornire una stima di quanto le varie squadre partecipanti al campionato NBA siano pericolose dal punto di vista offensivo.

3.5.1 Valutazione globale dei giocatori

Per prima cosa vengono quindi analizzati i risultati ottenuti a livello globale, senza fare distinzione tra i giocatori a livello di ruolo. Così facendo, si ottengono i risultati mostrati nella Tabella 3.3, nella quale sono stati inseriti solamente i 10 giocatori che hanno ottenuto le valutazioni più alte. Nell'Appendice B verrà poi riportata la tabella contenente anche la classifica completa.

Analizzando i giocatori presenti nella Tabella 3.3, si nota che questi sono quasi gli stessi che comparivano anche nella Tabella 2.5 ottenuta dall'applicazione del modello logistico stimato con il metodo di riduzione della distorsione di Firth. L'unica eccezione è data dall'ingresso nella classifica di Trae Young, il quale dopo aver disputato un'ottima stagione con gli *Atlanta Hawks* viene nominato seconda miglior matricola del torneo, arrivando secondo nel premio di *Rookie of the Year*. Per le analisi sugli altri giocatori presenti nella classifica valgono invece le considerazioni fatte alla fine del Capitolo 2.

Dalla similitudine tra i risultati ottenuti con l'approccio presentato nel capitolo precedente, si può quindi affermare che nei metodi applicati in questo capitolo vi è una netta predominanza dei dati rispetto all'informazione descritta nelle distribuzioni a priori, facendo sì che la distribuzione a posteriori si concentri maggiormente intorno alla funzione di verosimiglianza. Le poche differenze osservate possono quindi essere attribuite alla componente di dinamicità introdotta correggendo le distribuzioni a priori di ogni modello con le distribuzioni a posteriori del modello precedente.

3.5.2 Valutazione dei giocatori stratificata per ruolo

Dopo aver valutato i risultati a livello globale, vengono ora presentati quelli stratificati in base al ruolo di appartenenza dei cestisti. Per ciascun ruolo, verranno quindi selezionati i 10 giocatori che hanno ottenuto il *ranking* corretto più alto (utilizzando la formula (2.3)). I risultati ottenuti con la stratificazione coincidono con quelli del Capitolo 2 per quanto riguarda i ruoli indicati come ala grande, centro, e ala piccola, e pertanto verranno ora presentate solamente le classifiche ottenute per i ruoli identificati come *playmaker* e guardia tiratrice. I risultati sono riportati nelle Tabelle 3.4, 3.5.

Analizzando le Tabelle 3.4 e 3.5, si nota che queste differiscono rispetto alle Tabelle 2.8 e 2.9 solamente per un giocatore ciascuna. In particolare, si ha che con il metodo presentato in questo capitolo, Mike Conley viene identificato come un giocatore più pericoloso rispetto a Damian Lillard nel ruolo di *playmaker*. Per quanto riguarda invece il ruolo di guardia tiratrice, si ha che Zach LaVine viene considerato un giocatore più pericoloso a livello offensivo rispetto a Donovan Mitchell.

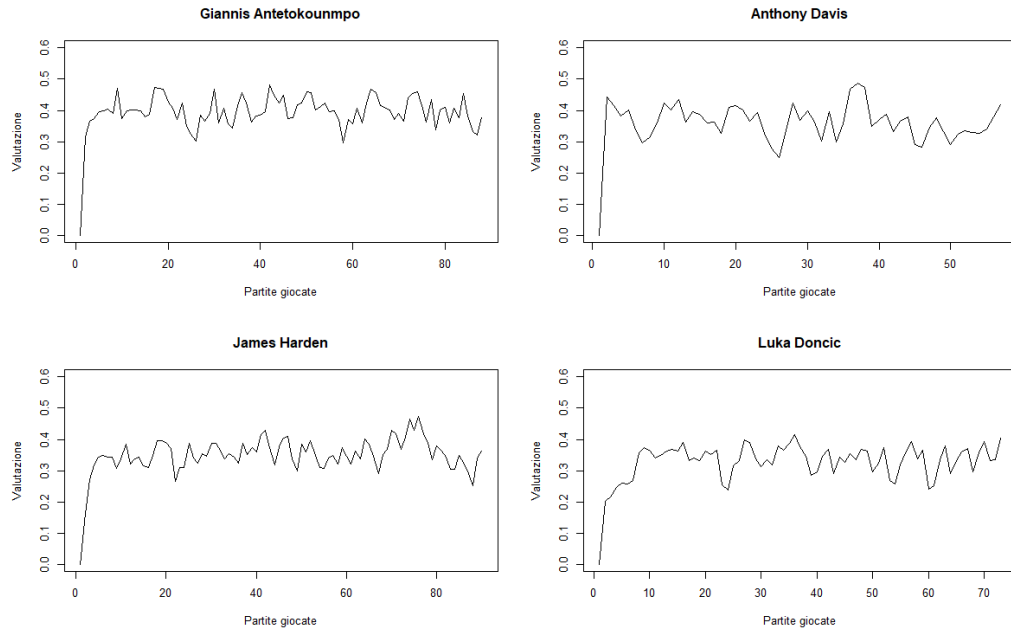
Tabella 3.4: Migliori 10 giocatori di ruolo playmaker classificati dal modello stimato con metodi Bayesiani

Giocatore	Squadra	Ranking	Ranking corretto
James Harden	HOU	0.36	22.57
Russell Westbrook	OKC	0.32	19.31
Trae Young	ATL	0.40	17.27
Stephen Curry	GSW	0.38	16.89
Kemba Walker	CHA	0.35	15.68
Mike Conley	MEM	0.38	15.14
Damian Lillard	POR	0.33	14.97
Ben Simmons	PHI	0.34	14.92
De'Aaron Fox	SAC	0.36	14.17
Lonzo Ball	LAL	0.46	14.16

Tabella 3.5: Migliori 10 giocatori di ruolo guardia tiratrice classificati dal modello stimato con metodi bayesiani

Giocatore	Squadra	Ranking	Ranking corretto
Luka Doncic	DAL	0.41	19.03
Jrue Holiday	NOP	0.39	17.46
Buddy Hield	SAC	0.49	16.80
Bradley Beal	WAS	0.36	16.26
DeMar DeRozan	SAS	0.36	15.73
Devin Booker	PHX	0.29	13.92
Zach LaVine	CHI	0.32	13.56
Donovan Mitchell	UTA	0.30	13.35
Klay Thompson	GSW	0.35	12.40
CJ McCollum	POR	0.33	11.87

Figura 3.5: Andamento delle performance stimate con metodi bayesiani dei migliori giocatori nel corso della stagione



La similitudine tra i risultati ottenuti in questi due capitoli sembra quindi portare ad una conferma della bontà delle analisi svolte. In Figura 3.5, vengono poi riportate le analisi delle performance per Giannis Antetokounmpo, James Harden, Anthony Davis e Luka Doncic, nel corso di tutta la stagione. Anche in questa figura, si nota una somiglianza nell'andamento delle performance dei giocatori rispetto a quelle ottenute con i metodi presentati nel Capitolo 2.

3.5.3 Valutazione delle performance di squadra

Infine, è stato deciso di proseguire le analisi con lo scopo di capire quali squadre sono quelle maggiormente orientate alla fase offensiva. Questo è stato fatto utilizzando la formula (2.4), la quale sfrutta le valutazioni sui giocatori ottenute con i metodi presentati in questo capitolo. Come nel capitolo precedente, la formula menzionata è stata applicata basandosi sulle

Tabella 3.6: Classifica delle performance offensive delle squadre NBA ottenute con metodi bayesiani

Squadra	Ranking	Squadra	Ranking
OKC	13.78	DEN	10.28
GSW	13.21	POR	10.10
HOU	12.54	ATL	10.00
NOP	11.74	MIA	9.98
MIL	11.61	BOS	9.88
UTA	11.24	MIN	9.70
DET	11.05	SAS	9.44
DAL	10.87	LAC	9.39
CLE	10.86	CHA	9.36
PHI	10.85	IND	9.17
WAS	10.61	MEM	8.81
CHI	10.58	PHX	8.80
SAC	10.52	ORL	8.64
LAL	10.51	BKN	8.56
TOR	10.45	NYK	7.89

valutazioni dei 200 giocatori con il *ranking* individuale più alto, portando ai risultati mostrati in Tabella 3.6.

Contrariamente ai risultati ottenuti per i giocatori, dove non si era verificata una distinzione particolarmente importante tra i metodi presentati nei due capitoli, nel caso della valutazione delle squadre si arriva a risultati che presentano alcune differenze rilevanti.

In particolare, le squadre che vengono classificate come le più pericolose rimangono sempre gli *Oklahoma Thunder* e i *Golden State Warriors*, così come quelle identificate come meno pericolose rimangono i *New York Knicks* e i *Brooklyn Nets*.

Per quanto riguarda le posizioni centrali della classifica si verificano invece dei cambiamenti. Più precisamente, si nota ad esempio che i *Milwaukee Bucks* scendono di due posizioni, a favore dei *New Orleans Pelicans* e degli *Houston*

Rockets, che con il loro gioco hanno contribuito a far diventare James Harden il miglior marcatore della stagione.

In conclusione, si può affermare che i risultati ottenuti a livello di squadra sembrano comunque soddisfacenti dopo averli confrontati con altre classifiche formulate da esperti del settore. Tuttavia, alcune valutazioni, come ad esempio la posizione dei *Toronto Raptors*, portano a affermare che la classifica delle squadre potrebbe essere migliorata approfondendo alcuni aspetti dello studio, come ad esempio l'introduzione di un effetto relativo alla squadra.

3.6 Un modello gerarchico con effetti annidati

Viste le ultime conclusioni tratte riguardo la stima delle performance di squadra, sembra quindi sensato considerare anche altri aspetti legati appunto alle squadre nella stima dei modelli. Per questo motivo e per provare a migliorare i risultati, è stato quindi deciso di utilizzare un modello gerarchico *bayesiano*, il quale rappresenta un'alternativa al modello con gli effetti fissi.

Per giustificare questa operazione, si può affermare che la modellazione gerarchica *bayesiana* è diventata un potente strumento in svariate tipologie di analisi grazie alla sua flessibilità e adattabilità nel trattare contesti complessi che coinvolgono sottogruppi, consentendo talvolta anche di aumentare l'efficienza nella stima. L'implementazione di questi modelli è dunque particolarmente sensata nel caso della presente tesi e permette di ottenere alcuni vantaggi importanti. Il primo di questi, è dato dal fatto che consente di introdurre un effetto di *borrowing strength*, ovvero consente di utilizzare le informazioni relative alla distribuzione di un determinato fattore, per fornire delle informazioni sulle distribuzioni di altri fattori (per altri dettagli si veda ad esempio Xu *et al.*, 2019). Inoltre, questo modello consente anche di essere esteso al caso di un modello con effetti annidati, dove due effetti, ed in particolare due fattori, si dicono annidati quando ogni livello di uno si incrocia con un solo livello dell'altro. Questi vantaggi permettono quindi di effettuare delle valutazioni più precise che comprendono oltre alle informazioni sui giocatori, anche quelle delle loro squadre di appartenenza.

Per le considerazioni fatte, è stato quindi adattato un modello gerarchico con effetti annidati (per approfondimenti si veda ad esempio Gelman e

Hill, 2007, Paragrafo 11), il quale contiene sia un effetto casuale di squadra che un effetto casuale di giocatore, annidato dentro a quello della squadra. L’inserimento di questi effetti porta a modificare il predittore lineare relativo all’azione i svolta dal giocatore j appartenente alla squadra s come

$$\eta_{ijs} = \text{squadra}_s \alpha_s + \text{giocatore}_j \alpha_{j[s]} + \mathbf{x}_i \beta,$$

dove α_s corrisponde all’effetto casuale relativo alla squadra e $\alpha_{j[s]}$ rappresenta l’effetto casuale del giocatore j annidato all’interno dell’effetto della squadra s .

Per fornire delle valutazioni sulle performance offensive dei giocatori, il modello è stato quindi stimato in ogni giornata in cui è stata disputata almeno una partita mediante la funzione `stan_glm` del pacchetto R `rstanarm` (Goodrich *et al.*, 2020). Questa, permette di fare inferenza *bayesiana* per modelli lineari generalizzati con coefficienti specifici di gruppo che hanno matrici di covarianza sconosciute, attraverso delle distribuzioni a priori flessibili. In particolare, per quanto riguarda i parametri di regressione non gerarchici sono state utilizzate come distribuzioni a priori delle t – *student* con 7 gradi di libertà. Per quanto riguarda invece le distribuzioni a priori dei parametri relativi agli effetti casuali, queste si basano su una decomposizione della matrice di covarianza (per ulteriori dettagli si veda ad esempio Gabry e Goodrich, 2020).

Dopodiché, utilizzando questo modello per prevedere le probabilità con cui un’azione viene indicata come pericolosa, è stato calcolato il *ranking* sia per i giocatori che per le squadre, utilizzando le formule (2.1), (2.2), (2.3) e (2.4).

L’applicazione delle formule menzionate ha consentito di ottenere i risultati mostrati nella Tabella 3.7 per quanto riguarda i giocatori, e nella Tabella 3.8 per quanto riguarda le squadre. Nella tabella relativa ai cestisti sono stati inseriti per questioni di spazio, solamente i 10 giocatori per i quali è stato stimato il *ranking* più alto. Nell’Appendice B vengono poi riportate le tabelle dove è stata effettuata una stratificazione per ruolo, che però ha portato a risultati simili rispetto a quelle presentate in precedenza, senza introdurre variazioni particolari nelle prime posizioni delle classifiche.

Tabella 3.7: Classifica dei 10 migliori giocatori NBA selezionati dal modello gerarchico stimato con il metodo bayesiano

Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
Giannis Antetokounmpo	MIL	PF	0.38	22.83
James Harden	HOU	PG	0.36	22.46
LeBron James	LAL	SF	0.38	20.49
Russell Westbrook	OKC	PG	0.33	19.90
Joel Embiid	PHI	C	0.33	19.83
Nikola Jokic	DEN	C	0.39	19.65
Anthony Davis	NOP	C	0.37	19.56
Paul George	OKC	SF	0.37	18.70
Luka Doncic	DAL	SG	0.40	18.59
Devin Booker	PHX	SG	0.37	17.79

Dalla Tabella 3.7, si notano alcune differenze rispetto alle classifiche presentate in precedenza. In particolare, si nota ad esempio che per Anthony Davis, giocatore che fino ad ora era sempre stato considerato tra i migliori 3, sono state stimate valutazioni inferiori, portandolo a scendere nella classifica generale, a favore di LeBron James. Interessante è però l’inserimento nella classifica di Paul George, il quale ha disputato una stagione eccellente, venendo anche candidato al premio di MVP. Andando poi a valutare l’andamento delle performance di alcuni giocatori nell’arco di tutta la stagione, si nota che la stima delle valutazioni ottenute utilizzando il modello gerarchico ha introdotto un effetto di *shrinkage*, come evidenziato dalla Figura 3.6. Da quest’ultima si nota infatti che rispetto ai metodi precedenti, l’utilizzo di effetti casuali annidati ha portato a comprimere maggiormente le valutazioni.

Andando ora ad analizzare i risultati a livello di squadra, i quali sono basati sulle valutazioni dei singoli giocatori, si nota che l’inserimento di un effetto casuale per le squadre ha portato ad alcuni cambiamenti. In particolare si nota che rispetto alle classifiche precedenti, a squadre che hanno disputato un’ottima stagione mostrando un buon gioco, come ad esempio i *Milwaukee Bucks* e i *Toronto Raptors*, sono state date delle valutazioni più

Figura 3.6: Andamento delle performance stimate con il modello gerarchico dei migliori giocatori nel corso della stagione

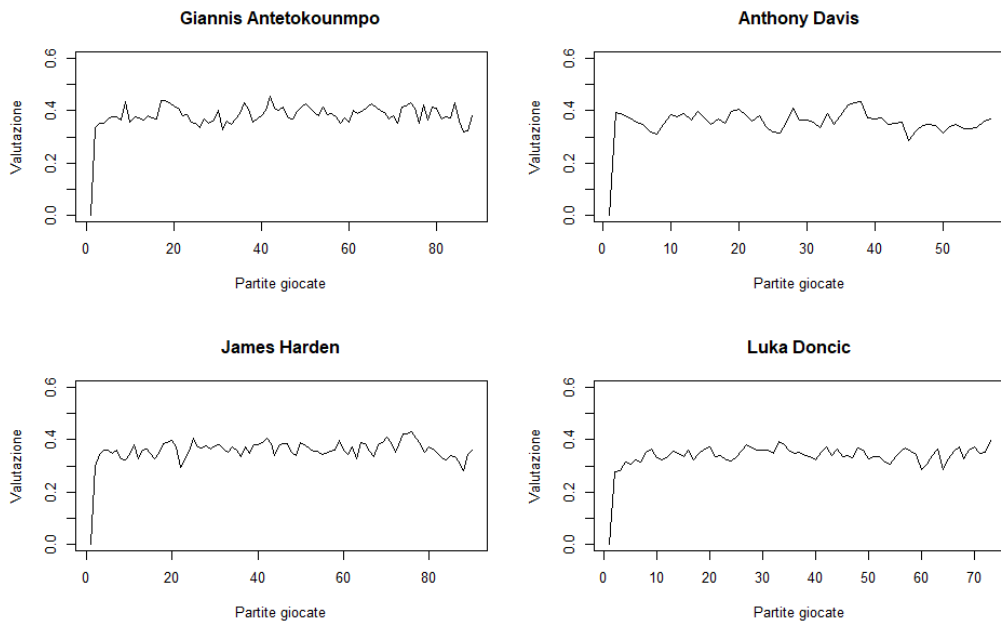


Tabella 3.8: Classifica delle performance offensive delle squadre NBA ottenute con il modello gerarchico stimato con metodi bayesiani

Squadra	<i>Ranking</i>	Squadra	<i>Ranking</i>
OKC	13.96	CHI	10.34
DET	12.60	LAL	10.26
GSW	12.10	DAL	9.99
MIL	11.65	LAC	9.89
HOU	11.59	IND	9.84
DEN	11.52	CLE	9.82
TOR	11.26	ATL	9.74
UTA	11.26	PHX	9.73
PHI	11.24	BOS	9.62
WAS	11.10	MEM	9.56
NOP	11.07	ORL	9.48
POR	11.00	MIA	9.26
SAC	10.99	CHA	9.06
MIN	10.66	BKN	8.61
SAS	10.48	NYK	8.09

alte. Anche in questo caso tuttavia, si ha che alcune valutazioni, come ad esempio il secondo posto nella classifica per i *Detroit Pistons*, non sembrano particolarmente sensate. Questo aspetto potrebbe essere dovuto al fatto che l'obiettivo primario delle analisi è sempre stato quello di valutare le performance dei singoli giocatori, o ancora al fatto che potrebbero essere utilizzate delle distribuzioni a priori maggiormente informative, dettate dalle conoscenze di esperti, o infine al fatto che questi indicatori considerano principalmente l'aspetto offensivo e non quello difensivo, altrettanto importante per una squadra vincente.

3.7 Valutazioni finali

In conclusione, l'applicazione di un modello di regressione logistica stimato con un metodo *bayesiano* ha permesso di fornire una valutazione delle performance offensive dei giocatori che sembra essere soddisfacente da un punto di vista qualitativo. Confrontando infatti i risultati con altre classifiche costruite da esperti e con la classifica stilata nel Capitolo 2, si evince che i dati portano a considerazioni simili nonostante vengano utilizzati approcci diversi, portando quindi a valutare positivamente i risultati.

L'applicazione delle tecniche statistiche utilizzate in questo capitolo, comporta inoltre alcuni vantaggi, che sebbene non siano particolarmente evidenti nel caso della presente tesi, potrebbero risultare importanti in altri studi. In particolare, l'introduzione delle correzioni delle distribuzioni a priori ha infatti permesso di introdurre un'ulteriore componente di dinamicità nel modello, consentendo di tenere maggiormente conto dello stato di forma dei giocatori. Infine, l'utilizzo di un modello gerarchico ha permesso di considerare in forma maggiore le caratteristiche delle squadre, velocizzando le operazioni di stima.

Per via di questi aspetti, sembra dunque che il metodo presentato in questo capitolo consenta di affrontare il problema di valutare le performance dei cestisti fornendo risultati affidabili. Un possibile miglioramento, che potrebbe essere oggetto di studi futuri, è tuttavia quello di utilizzare delle distribuzioni a priori maggiormente informative, dettate dalla conoscenza di esperti del settore. In questo modo si andrebbe quindi a convogliare l'informazione

presente nei dati con quella di alcuni esperti, e ciò potrebbe far migliorare ulteriormente i risultati.

Conclusioni

Nella presente tesi sono stati proposti dei metodi per la valutazione delle performance offensive dei giocatori di basket utilizzando i dati relativi al campionato NBA disputatosi negli anni 2018 e 2019. In particolare, dopo aver presentato un metodo basato sui dati per definire se ciascuna azione compiuta dai giocatori può essere considerata importante dal punto di vista offensivo, sono state applicate diverse tecniche inferenziali per stimare le probabilità che ogni azione effettuata risulti pericolosa. Dopodiché, sfruttando le probabilità stimate, sono stati valutati i giocatori nel corso di tutta la stagione, consentendo di analizzare l'andamento a livello prestazionale di ciascun cestista NBA e permettendo quindi di formulare dei confronti mirati tra questi.

Nel dettaglio, per stimare le valutazioni dei giocatori sono stati presentati principalmente due metodi. Il primo, che prevede di utilizzare per la stima delle probabilità di pericolosità delle azioni un modello di regressione logistica stimato con il metodo della riduzione della distorsione proposto da Firth. Il secondo invece, che prevede di stimare le probabilità con cui ogni azione viene considerata pericolosa utilizzando un modello logistico stimato con delle tecniche *bayesiane*. Per entrambi i metodi, per fornire delle valutazioni sui giocatori tenendo in considerazione l'andamento temporale delle loro prestazioni, è stato poi deciso di aggregare i risultati forniti dai modelli stimati in ciascuna giornata utilizzando la statistica EWMA.

Un'aspetto particolarmente interessante che contraddistingue il primo metodo presentato da quello *bayesiano*, è dato dal fatto che in quest'ultimo è stato possibile introdurre anche un'ulteriore componente dinamica nella valutazione dei giocatori, inserendo nei modelli stimati in ogni giornata delle

informazioni a priori sui parametri, derivanti dalla stima dei modelli ottenuta nelle giornate precedenti. Con questo approccio, è poi stato possibile includere nei modelli anche degli effetti casuali annidati, che corrispondono ad un effetto casuale di squadra e ad uno di giocatore, annidato dentro a quello della squadra. Ciò ha consentito di includere più informazione nel modello, migliorando i risultati dal punto di vista interpretativo.

In conclusione, dopo aver confrontato le valutazioni fornite dalle due procedure presentate con quelle di alcuni esperti del settore, sembra che i metodi proposti forniscano risultati soddisfacenti. Queste tecniche consentono infatti di fornire delle valutazioni complete sotto vari aspetti, quali l'analisi dei singoli giocatori, delle singole squadre, o della tipologia di giocate effettuate. Ad esempio, consentono sia di confrontare i giocatori in base al ruolo di appartenenza, che di analizzare tutte le azioni nel dettaglio, permettendo anche di comprendere quali sono le giocate che maggiormente risultano incisive in determinate partite. Queste analisi, mettendo a disposizione un approccio basato sui dati che utilizza anche dei concetti probabilistici, portano quindi a valutazioni di vario genere, che potrebbero risultare utili alle squadre NBA, consentendo di integrare ai semplici indici descrittivi anche altre considerazioni più approfondite.

A livello di risultati ottenuti, si è poi osservato che le due metodologie utilizzate hanno portato in alcune situazioni a risultati molto simili, ad evidenziare come i dati predominino sul fenomeno studiato. Questo aspetto, porta dunque ad una ulteriore conferma della bontà delle valutazioni ottenute, introducendo però anche un possibile miglioramento. Infatti, se le tecniche *bayesiane* portano a risultati simili rispetto al modello logistico stimato con la correzione per la distorsione di Firth, si può affermare che le distribuzioni a priori utilizzate non influiscano particolarmente nella stima. Per questo motivo, sembra dunque una buona idea approfondire questo aspetto, con lo scopo di selezionare delle distribuzioni a priori maggiormente informative, che potrebbero ad esempio essere ricavate dalle valutazioni di esperti del settore.

Per quanto riguarda le criticità dei metodi utilizzati invece, la più rilevante è quella relativa ai tempi computazionali necessari per fornire le valutazioni. Questo problema, che è stato parzialmente risolto con l'utilizzo della statistica EWMA, è più evidente quando si utilizza un approccio di tipo *bayesiano*.

Queste tecniche inferenziali infatti, pur permettendo di considerare maggiormente la dinamicità del problema, hanno un carico computazionale di gran lunga maggiore rispetto al metodo proposto nel Capitolo 2, poiché per ogni giornata in cui viene stimato un modello bisogna utilizzare l'*Hamiltonian Monte Carlo*.

Un altro dei problemi osservati nella valutazione dei cestisti è dato poi dall'alta dimensionalità del problema. Avendo infatti un livello di dettaglio particolarmente alto per la descrizione delle azioni, e utilizzando nella stima dei modelli degli effetti fissi e/o casuali per i giocatori, si ha che il numero di parametri da stimare è molto elevato. Questo aspetto, aggrava ulteriormente i problemi computazionali, e in alcuni casi introduce una quasi perfetta separazione tra le variabili, causando dei problemi di stima. Per questo motivo, potrebbe quindi risultare interessante valutare in futuro dei metodi che consentano di fare selezione delle variabili, così da risolvere almeno in parte questo problema.

Dalle analisi svolte, sembra infine sensato presentare alcune idee, che se approfondite in futuro potrebbero consentire di migliorare i risultati ottenuti, permettendo delle valutazioni più approfondite e complete. Per prima cosa, vista la volontà di valutare i giocatori sulla base delle azioni che questi hanno svolto, potrebbe essere particolarmente utile introdurre nella raccolta dati anche la posizione dei giocatori. Viste le moderne capacità di raccolta, si potrebbe infatti rilevare la posizione dei cestisti sia al momento iniziale dell'azione, che anche nel momento in cui questi si liberano del pallone. Considerando anche questo aspetto, sarebbe quindi possibile fornire delle valutazioni più precise ed esaustive.

Infine, un ulteriore aspetto che potrebbe essere interessante da approfondire in futuro, riguarda l'analisi delle performance di squadra. Nella presente tesi infatti, essendo che l'obiettivo era principalmente quello di valutare i giocatori dal punto di vista offensivo, questo aspetto ha assunto un ruolo secondario. Tuttavia, sfruttando i dati presentati nel Capitolo 1, si potrebbero fornire risultati particolarmente utili anche in questo ambito. Disponendo infatti di un *dataset play by play*, si potrebbero utilizzare altri strumenti statistici come ad esempio le reti, per valutare anche la tipologia di gioco delle squadre, e fornire perciò risultati utili agli allenatori delle squadre NBA per

Stima *bayesiana* del modello logistico per la valutazione dei giocatori NBA

preparare le partite da disputare.

Bibliografia

- Azzalini, Adelchi e Bruno Scarpa (2004). *Analisi dei dati e data mining*. Springer-Verlag.
- Basketball-reference (2019). *Basketball Stats and History*. URL: <https://www.basketball-reference.com/>.
- Betancourt, Michael (2017). *How the Shape of a Weakly Informative Prior Affects Inferences*. URL: https://mc-stan.org/users/documentation/case-studies/weakly_informative_shapes.html.
- BigDataBall (2007). *Dataset NBA 2018/2019*. URL: <https://www.bigdataball.com/datasets/nba/play-by-play/>.
- Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li e Allen Riddell (2017). «Stan: A Probabilistic Programming Language». In: *Journal of Statistical Software* **76**, 1–32.
- Chib, Siddhartha e Edward Greenberg (1995). «Understanding the Metropolis-Hastings Algorithm». In: *The American Statistician* **49**, 327–335.
- Dandolo, David (2019). «Valutazione statistica delle performance dei giocatori della Serie A». Tesi di laurea magistrale. Università degli studi di Padova.
- Decroos, Tom (2019). «Actions Speak Louder than Goals: Valuing Player Actions in Soccer». In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- Duane, Simon, A.D. Kennedy, Brian J. Pendleton e Duncan Roweth (1987). «Hybrid Monte Carlo». In: *Physics Letters B* **195**, 216–222.
- Efron, Bradley (2013). «Bayes' Theorem in the 21st Century». In: *Science* **340**, 1177–1178.

BIBLIOGRAFIA

- ESPN (2019). *NBA rank top 10*. URL: https://www.espn.com/nba/story/_/id/24668720/nbarank-2018-19-1-10-best-players-season.
- Firth, David (1993). «Bias Reduction of Maximum Likelihood Estimates». In: *Biometrika* **80**, 27–38.
- Gabry, Jonah e Ben Goodrich (2020). *Estimating Generalized (Non-)Linear Models with Group-Specific Terms with rstanarm*. URL: <http://mc-stan.org/rstanarm/articles/glmer.html>.
- Gamerman, Dani e Hedibert F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman e Hall.
- Gelman, Andrew, W. R. Gilks e Gareth Roberts (1997). «Weak convergence and optimal scaling of random walk Metropolis algorithms». In: *Annals of Applied Probability* **7**, 110–120.
- Gelman, Andrew e Jennifer Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gelman, Andrew e Donald Rubin (1992). «Inference from iterative simulation using Multiple Sequences». In: *Statistical Science* **7**, 457–511.
- Girolami, Mark e Ben Calderhead (2011). «Riemann manifold Langevin and Hamiltonian Monte Carlo methods». In: *Journal of the Royal Statistical Society: Series B (Statistical methodology)* **73**, 123–214.
- Goodrich, Ben, Jonah Gabry, Imad Ali e Sam Brilleman (2020). *rstanarm: Bayesian applied regression modeling via Stan*. R package version 2.21.1. URL: <https://mc-stan.org/rstanarm>.
- Heeren, Dave (1988). *Basketball Abstracts*. Prentice Hall.
- Hoffman, Matthew e Andrew Gelman (2014). «The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo». In: *Journal of Machine Learning Research* **15**, 1351–1381.
- Hunter, Stuart (1986). «The Exponentially Weighted Moving Average». In: *Journal of Quality Technology* **18**, 203–210.
- IBM (1945). *Most Valuable Player IBM*. URL: <https://www.research.ibm.com/>.
- Kay, H. (1966). «A Statistical Analysis Of The Profile Technique For The Evaluation Of Competitive Basketball Performance». Master's Thesis. University Of Alberta.

-
- Kim, Junhyung Lyle, Dustin Tran e Panos Toulis (2019). *sgd: Stochastic Gradient Descent for Scalable Estimation*. R package version 1.1.1. URL: <https://CRAN.R-project.org/package=sgd>.
- Kosmidis, Ioannis e David Firth (2020). «Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models». In: *Biometrika*. In stampa.
- Kosmidis, Ioannis, Euloge Clovis Kenne Pagui e Nicola Sartori (2020). «Mean and median bias reduction in generalized linear models». In: *Statistics and Computing* **30**, 43–59.
- McCullagh e Nelder (1989). *Generalized Linear Models*. Chapman e Hall.
- NBAstuffer (2007). *Basketball Research*. URL: <https://www.nbastuffer.com/>.
- Neal, Radford (2011). *Handbook of Markov Chain Monte Carlo*. Chapman e Hall.
- Oliver, Dean (2004). *Basketball on Paper: Rules and Tools for Performance Analysis*. Potomac Books Inc.
- Pace, Luigi e Alessandra Salvan (1996). *Teoria della statistica. Metodi, modelli, approssimazioni asintotiche*. CEDAM.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Rezzani, Alessandro (2013). *Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*. Apogeo Education - Maggioli Editore.
- Robert, Christian (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer.
- Salvan, Alessandra, Nicola Sartori e Luigi Pace (2020). *Modelli lineari generalizzati*. Springer-Verlag.
- Simovic, Slobodan, Bojan Matkovic, Mihajlo Mijanovic e Milenko Vojvodic (2019). «Comparative values, correlation and classification of basketball players based on the efficiency index and expert evaluation by coaches». In: *International Journal of Physical Education, Fitness and Sports* **8**, 76–87.

BIBLIOGRAFIA

- Wickham, Hadley, Romain François, Lionel Henry e Kirill Müller (2020). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.0. URL: <https://CRAN.R-project.org/package=dplyr>.
- Xu, Ganggang, Huirong Zhu e Jack Lee (2019). «Borrowing strength and borrowing index for Bayesian hierarchical models». In: *Computational Statistics and Data Analysis* **144**, 106901.

Appendice A

Codice R utilizzato

In questa Appendice viene riportato il codice R che è stato utilizzato per svolgere le analisi.

Nel seguito vi è il codice implementato nel Capitolo 1 per svolgere alcune delle operazioni di pulizia e per la creazione della variabile risposta.

Codice A.1: Librerie utilizzate nella fase di costruzione del *dataset*

```
library(dplyr)
library(fastDummies)
library(stringr)
library(stringi)
library(tidyr)
library(readxl)
```

Codice A.2: Operazioni di pulizia

```
setwd("D:/Esplora/Documents/Tesi/2018-2019_NBA_PbP_Logs")
Nba_17_18 = read.csv("[10-16-2018]-[06-13-2019]-combined-stats.csv")
str(Nba_17_18)
summary(Nba_17_18)

# Rimozione di tutte le righe che non sono associate a nessun giocatore
Nba_17_18 = Nba_17_18[Nba_17_18$player != "", ]
```

```
errori = NULL
for (i in 1:nrow(Nba_17_18)) {
  if((Nba_17_18$player[i] %in% c(Nba_17_18$a1[i],Nba_17_18$a2[i],Nba
    _17_18$a3[i],Nba_17_18$a4[i],Nba_17_18$a5[i],Nba_17_18$h1[i],
    Nba_17_18$h2[i],Nba_17_18$h3[i],Nba_17_18$h4[i],Nba_17_18$h5[i
    ],Nba_17_18$left[i])) == FALSE)
    errori = c(errori,i)
}
Nba_17_18 = Nba_17_18[-errori,]

# Creazione delle variabili dummy per gli eventi e i sotto-eventi
evento = dummy_cols(Nba_17_18$event_type,remove_first_dummy = T)
evento = evento[,-1]
drop_column = (apply(evento,2,sum) != 0)
evento = evento[,drop_column]
nomi_evento = c("foul","free_throw","jump_ball","miss","rebound","shot","
  sub","turnover","violation")
colnames(evento) = nomi_evento
drop.evento = which(colnames(evento)=="sub")
evento = evento[,-drop.evento]
nomi_subevento = as.character(unique(Nba_17_18$type))
length(nomi_subevento)
idx_free_throw = grep("Free",nomi_subevento)
idx_violation = grep("violation",nomi_subevento)
idx_foul = grep("foul",nomi_subevento)
idx_ejection = grep("Ejection",nomi_subevento)
idx_sub = grep("sub",nomi_subevento)
idx = c(idx_free_throw,idx_ejection,idx_foul,idx_violation,idx_sub)
Nba_17_18$type[(Nba_17_18$type %in% nomi_subevento[idx])] = "Altro"
sub_evento = dummy_cols(Nba_17_18$type,remove_first_dummy = T)
sub_evento = sub_evento[,-1]
drop_column = (apply(sub_evento,2,sum) > 50)
sub_evento = sub_evento[,drop_column]
nomi_subevento = colnames(sub_evento)
nomi_subevento = str_remove_all(nomi_subevento,".*\\_")
nomi_subevento = str_replace_all(nomi_subevento," ","_")
```

```

colnames(sub_evento) = nomi_subevento
sub_evento = sub_evento[, -which(colnames(sub_evento) == "Altro")]

# Creazione della matrice delle variabili esplicative
dati = Nba_17_18[, c("date", "game_id", "data_set", "player", "team", "period", "
  elapsed", "play_length", "home_score", "away_score", "result", "shot_
  distance", "num", "outof", "points", "original_x", "original_y", "converted_
  x", "converted_y", "assist", "block", "steal", "description")]
dati = cbind(dati, evento)
dati = cbind(dati, sub_evento)

# Inserimento della riga relativa al giocatore che effettua l'assist
dati$elapsed = as.difftime(dati$elapsed)
n_assist = nrow(dati[dati$assist != "",])
vett = matrix(NA, nrow = n_assist, ncol = 77)
colnames(vett) = c("idx", colnames(dati)[-20], "assist")
j = 1
for(i in 1:nrow(dati)) {
  if(dati$assist[i] != "") {
    vett[j,] = as.array(unlist(c(i, dati[i, 1:3], dati$assist[i],
      dati[i, 5:6], dati$elapsed[i]-1, "0:00:01", dati$home_
      score[i-1], dati$away_score[i-1], "", rep(NA, 4), rep("", 7)
      , rep(0, 53), 1)))
    j = j+1
  }
}
row.names(dati) = 1:nrow(dati)
dati = dati[, -c(which(colnames(dati) == "assist"))]
vett = as.data.frame(vett)
row.names(vett) = vett$idx
vett = vett[, -1]
dati$assist = rep(0, nrow(dati))
dati$idx = 1:nrow(dati)
vett$idx = as.numeric(row.names(vett))
dati = rbind(dati, vett)
dati = dati[order(dati$idx, dati$period, dati$elapsed),]

```

```

dati = dati[,-77]

# Inserimento della riga relativa al giocatore che ruba la palla
dati$elapsed = as.numeric(dati$elapsed)
n_steal = nrow(dati[dati$steal != "",])
vett = matrix(NA,nrow = n_steal,ncol = 77)
colnames(vett) = c("idx",colnames(dati)[-21],"steal")
j = 1
for(i in 1:nrow(dati)) {
  if(dati$steal[i] != "") {
    vett[j,] = as.array(unlist(c(i,dati[i,1:3],dati$steal[i],
      player_team$team[player_team$player == dati$steal[i] &
      player_team$game_id == dati$game_id[i]][1],dati[i,6],
      dati$elapsed[i]+1,"0:00:01",dati$home_score[i-1],dati$
      away_score[i-1],"",rep(NA,4),rep("",7),rep(0,53),1)))
    j = j+1
  }
}
row.names(dati) = 1:nrow(dati)
dati = dati[,-c(which(colnames(dati) == "steal"))]
vett = as.data.frame(vett)
row.names(vett) = vett$idx
vett = vett[,-1]
dati$steal = rep(0,nrow(dati))
dati$idx = 1:nrow(dati)
vett$idx = as.numeric(row.names(vett))
dati = rbind(dati,vett)
dati = dati[order(dati$idx,dati$period,dati$elapsed),]
dati = dati[,-77]

# Inserimento della riga relativa al giocatore che effettua una stoppata
dati$elapsed = as.numeric(dati$elapsed)
n_block = nrow(dati[dati$block != "",])
vett = matrix(NA,nrow = n_block,ncol = 77)
colnames(vett) = c("idx",colnames(dati)[-20],"block")
j = 1

```

```

for(i in 1:nrow(dati)) {
  if(dati$block[i] != "") {
    vett[j,] = as.array(unlist(c(i,dati[i,1:3],dati$block[i],
      player_team$team[player_team$player == dati$block[i] &
      player_team$game_id == dati$game_id[i]][1],dati[i,6],
      dati$elapsed[i]+1,"0:00:01",dati$home_score[i-1],dati$
      away_score[i-1], "", rep(NA,4), rep("",7), rep(0,53),1)))
    j = j+1
  }
}
row.names(dati) = 1:nrow(dati)
dati = dati[,-c(which(colnames(dati) == "block"))]
vett = as.data.frame(vett)
row.names(vett) = vett$id
vett = vett[,-1]
dati$block = rep(0,nrow(dati))
dati$id = 1:nrow(dati)
vett$id = as.numeric(row.names(vett))
dati = rbind(dati,vett)
dati = dati[order(dati$id,dati$period,dati$elapsed),]
dati = dati[,-77]

# Accorpamento delle colonne miss e shot, poiche si riferiscono entrambe
  al tiro a canestro
dati$shot = as.numeric(dati$shot)
dati$miss = as.numeric(dati$miss)
dati$shot = dati$shot+dati$miss
dati = dati[,-which(colnames(dati) == "miss")]

# Rimozione giocatori non assegnati a nessuna squadra
dati = dati[-which(is.na(dati$team)),]

# Rimozione giocatori con poche azioni
d = dati %>%
group_by(player) %>%
summarise(numAzioni = n())

```

```
drop_player = d$player[d$numAzioni > 100]
dati = dati[dati$player %in% drop_player,]
```

Codice A.3: Creazione della variabile risposta

```
y = NULL
k_lag = 4
game = unique(dati$game_id)
for(i in 1:length(game)) {
  cat(i, " ")
  k = k_lag
  dati_game = dati[dati$game_id == game[i],]
  squadre = unique(dati_game$team)
  for(j in 1:nrow(dati_game)) {
    counter = 0
    null_counter = 0
    squadra_avv = squadre[-which(squadre %in% dati_game$team[j
    ])]
    if(j+k > nrow(dati_game))
      k = k - 1
    for(h in 0:k) {
      if(dati_game$team[j] == dati_game$team[j+h] & dati
      _game$result[j+h] == "made")
        counter = counter + 1
      if(squadra_avv == dati_game$team[j+h] & dati_game$
      result[j+h] == "made" )
        null_counter = 1
    }
    if(counter > 0 & null_counter == 0)
      y = c(y,1)
    else
      y = c(y,0)
  }
}
dati$y = y
```

Ora viene riportato il codice utilizzato nel Capitolo 2. In particolare, sarà presente il codice per la stima del modello logistico e quello per confrontare il modello contenente solo le variabili relative alle azioni con quello che contiene anche gli effetti fissi dei giocatori. Successivamente verrà presentata la parte dove è stato applicato di partita in partita il modello logistico stimato con il metodo di riduzione della distorsione proposto da Firth. Infine viene riportato il codice per la costruzione del *ranking* sfruttando le proprietà della statistica EWMA.

Codice A.4: Librerie utilizzate nelle analisi relative al Capitolo 2

```
library(dplyr)
library(brglm2)
library(ROSE)
library(Matrix)
library(sgd)
library(biglm)
```

Codice A.5: Confronto tra il modello contenente solo le variabili relative alle azioni con quello che contiene anche gli effetti fissi dei giocatori

```
# Aumento della memoria disponibile: altrimenti non si riesce a stimare il
  modello
memory.limit(48846)
# Modello senza effetti fissi
drop.column = c(1:14, which(colnames(dati) == "unknown"), which(colnames(
  dati) == "jump_ball.1"), which(colnames(dati) == "rebound"), which(
  colnames(dati) == "turnover.1"))
m.logistico = glm(y ~ ., family = binomial, data = dati[, -drop.column])
# Modello con effetti fissi
drop.column1 = c(1:3, 5:14, which(colnames(dati) == "unknown"), which(
  colnames(dati) == "jump_ball.1"), which(colnames(dati) == "rebound"),
  which(colnames(dati) == "turnover.1"))
nba.formula = as.formula(paste("y ~ -1 +", paste(colnames(dati[, -c(drop.
  column1, 70, 73)]), collapse = "+")))
m.logistico1 = glm(nba.formula, family = binomial, data = dati)
# Confronto i due modelli tramite il test del rapporto di verosimiglianza
```

```
anova(m.logistico,m.logistico1,test = "LRT")
```

Codice A.6: Valutazione del modello dal punto di vista della classificazione

```
p.logistico1 = predict(m.logistico1,type = "response")
roc.curve(dati$y,as.vector(p.logistico1), main = "", xlab = "Tasso di
  falsi positivi", ylab = "Tasso di veri positivi")
accuracy.meas(dati$y,p.logistico1)
prev.logistico1 = ifelse(p.logistico1 > 0.5,1,0)
t.prev = table(dati$y,prev.logistico1)
```

Codice A.7: Stima del modello logistico giornata dopo giornata, con il metodo della riduzione della distorsione di Firth

```
game = unique(dati$date)
player.game = unique(dati$player)
# Matrice dove vengono salvate le valutazioni sui giocatori stimate
rank.match = matrix(rep(0,length(game)*length(player.game)), nrow = length
  (player.game), ncol = length(game))
row.names(rank.match) = NULL
colnames(rank.match) = 1:length(game)
# Stima del modello giornata dopo giornata
for(i in 1:length(game)) {
  # Rimozione giocatori che nel match considerato hanno partecipato
  a poche azioni
  stop_player = dati[dati$date == game[i],] %>%
  group_by(player) %>%
  summarize(numAzioni = n(),.groups = 'drop')
  stop_player = stop_player$player[stop_player$numAzioni > 5]
  x.game = dati[dati$date == game[i] & dati$player %in% stop_player
    ,-c(drop.column1,70,73)]
  y.game = dati[dati$date == game[i] & dati$player %in% stop_player
    ,70]
  # Rimozione di variabili che nel match considerato sono costanti
  c.remove = (apply(x.game[, -c(1,53)],2,sum) > 5)
  c.remove = c(T,c.remove[1:51],T,c.remove[52])
}
```

```

x.game = x.game[,c.remove]
# Stima del modello
m.game = glm(y.game ~ ., family = binomial("logit"), data = x.game
            ,method = "brglmFit", type = "MPL-Jeffreys")
# Previsioni per i giocatori con i coefficienti stimati
p.game.player = predict(m.game,type = "response")
prev.totali = data.frame(player = as.vector(x.game$player),
                        percentuale = p.game.player)
prev.totali = prev.totali %>%
group_by(player) %>%
summarise(RatingMatch = mean(percentuale),.groups = 'drop')
r.m = data.frame(player = unique(dati$player))
r.m = r.m %>%
left_join(prev.totali,by = c("player" = "player"))
r.m$RatingMatch[is.na(r.m$RatingMatch)] = 0
rank.match[,i] = r.m$RatingMatch
cat(i," ")
}

```

Codice A.8: Utilizzo della statistica EWMA per aggregare i risultati

```

lambda = 0.5
num.match = length(game)
num.giocatori = length(player.game)
ewma.rank = matrix(rep(0,num.match*num.giocatori),nrow = num.giocatori,
                  ncol = num.match)
for(i in 1:num.match) {
  x.game = dati[dati$date == game[i],]
  for(j in 1:num.giocatori) {
    if(nrow(x.game[x.game$player == player.game[j],]) == 0 & i
      != 1) {
      ewma.rank[j,i] = ewma.rank[j,i-1]
    }
    else if(nrow(x.game[x.game$player == player.game[j],]) >
      0) {
      p.game.player = rank.match[j,i]
    }
  }
}

```

```
        if(i == 1) {
            ewma.rank[j,i] = p.game.player
        }
        else if(ewma.rank[j,i-1] == 0) {
            ewma.rank[j,i] = p.game.player
        }
        else {
            ewma.rank[j,i] = (1-lambda)*ewma.rank[j,i
                -1]+lambda*p.game.player
        }
    }
}
cat(i, " ")
}
```

Codice A.9: Calcolo del *ranking* dei giocatori di basket sfruttando i risultati del modello stimato dopo l'ultima giornata

```
R = ewma.rank[,213]
ris = data.frame(player = player.game,ranking = R,row.names = 1:length(
    player.game))
ris = dati %>%
group_by(player,team,ruolo) %>%
summarise() %>%
right_join(ris) %>%
arrange(desc(ranking))
# Aggiustamento del ranking per il numero di azioni
d = dati %>%
group_by(player,team) %>%
summarise(numAzioni = n())
ris = ris %>%
left_join(d)
G = dati %>%
group_by(player,game_id) %>%
summarise()
G = G %>%
```

```

group_by(player) %>%
summarise(numPartite = n()) %>%
arrange(desc(numPartite))
# Calcolo finale del ranking corretto per il numero di azioni
ris = ris %>%
left_join(G)
ris$finalRank = ris$ranking*ris$numAzioni/ris$numPartite
ris = ris %>%
arrange(desc(finalRank))
ris

```

Codice A.10: Calcolo del *ranking* di squadra sfruttando i risultati del modello stimato dopo l'ultima giornata

```

# Valutazione della squadra
ris1 = ris[1:200,] %>%
group_by(Squadra) %>%
summarise(rankingSquad = mean(FR), .groups = 'drop')
ris1 = ris1 %>% arrange(desc(rankingSquad))
ris1

```

In conclusione, viene quindi riportato il codice utilizzato nelle analisi svolte nel Capitolo 3. In particolare, sarà presente il codice relativo alla stima dei modelli mediante metodi *bayesiani*, sia nel caso dove vengono utilizzate sempre le distribuzioni a priori di default che nel caso dove vengono utilizzate come distribuzioni a priori le posteriori dei modelli precedenti. Il codice qui riportato si riferisce al caso dove vengono inseriti nel modello gli effetti fissi per i cestisti. Successivamente viene anche riportato il codice che stima il modello utilizzando degli effetti casuali annidati: un effetto casuale di squadra e un effetto casuale di giocatore, annidato dentro a quello della squadra. Non verrà qui riportato il codice che aggrega i risultati dei modelli utilizzando la statistica EWMA e il codice per il calcolo del *ranking* poiché è analogo a quello utilizzato nel Capitolo precedente.

Codice A.11: Librerie utilizzate nelle analisi relative al Capitolo 3

```
library(dplyr)
library(stringr)
library(tidyr)
library(mvtnorm)
library(bayesplot)
library(brms)
library(rstanarm)
library(TeachingDemos)
```

Codice A.12: Stima del modello logistico con metodi bayesiani dopo ogni giornata, utilizzando le distribuzioni a priori di default

```
niter = 4000
warmup = 2000
game = unique(dati$date)
player.game = unique(dati$player)
num.match = length(game)
num.giocatori = length(player.game)
rank.match = matrix(rep(0, length(game)*length(player.game)), nrow = length(
  player.game), ncol = length(game))
row.names(rank.match) = NULL
colnames(rank.match) = 1:num.match
for(match_id in 1:num.match) {
  # Rimozione variabili costanti
  null.var = which(apply(X[dati$date == game[match_id],], 2, sum) ==
    0)
  constant.variable = which(colnames(dati) %in% names(null.var))
  nba.formula = as.formula(paste("y ~ -1 +", paste(colnames(dati[, -c
    (drop.column1, constant.variable, 70)]), collapse = "+")))
  x.game = dati[dati$date == game[match_id], -c(constant.variable, 70)
  ]
  # Utilizzo rstanarm per stimare il modello logistico
  mcmc.rstan = stan_glm(nba.formula, family = binomial(), data =
    dati[dati$date == game[match_id],], iter = niter, warmup =
    warmup, cores = 4, verbose = F)
  # Previsioni per i giocatori
```

```

    p.lin.pred = posterior_linpred(mcmc.rstan)
    pi.chain.action = plogis(p.lin.pred)
    p.game.player = apply(pi.chain.action,2,mean)
    prev.totali = data.frame(player = as.vector(dati[dati$date == game
      [match_id],"player"]),percentuale = p.game.player)
    prev.totali = prev.totali %>%
    group_by(player) %>%
    summarise(RatingMatch = mean(percentuale),.groups = 'drop')
    r.m = data.frame(player = unique(dati$player))
    r.m = r.m %>%
    left_join(prev.totali,by = c("player" = "player"))
    r.m$RatingMatch[is.na(r.m$RatingMatch)] = 0
    rank.match[,match_id] = r.m$RatingMatch
    cat(match_id," ")
}

```

Codice A.13: Stima del modello logistico con metodi bayesiani dopo ogni giornata, utilizzando le distribuzioni a priori corrette per le posteriori dei modelli precedenti

```

niter = 4000
warmup = 2000
game = unique(dati$date)
player.game = unique(dati$player)
num.match = length(game)
num.giocatori = length(player.game)
rank.match = matrix(rep(0,length(game)*length(player.game)), nrow = length
  (player.game), ncol = length(game))
row.names(rank.match) = NULL
colnames(rank.match) = 1:num.match
drop.column1 = c(1:3, 5:14, which(colnames(dati) == "unknown"), which(
  colnames(dati) == "jumpball.1"), which(colnames(dati) == "rebound"),
  which(colnames(dati) == "turnover.1"), 73)
# Nuova matrice del disegno e variabile risposta
X = model.matrix(as.formula(paste("~ -1 +", paste(colnames(dati[, -c(drop.
  column1,70)]),collapse = "+"))), data = dati)

```

```
y = dati[, "y"]
N = nrow(dati)
p = ncol(X)
# Matrici per l'aggiornamento delle distribuzioni a priori da utilizzare
coefficients.giornata = matrix(rep(0, num.match*p), nrow = num.match, ncol =
  p)
sd.giornata = matrix(rep(2.5, num.match*p), nrow = num.match, ncol = p)
lambda = 0.5
for(match_id in 1:num.match) {
  # Rimozione delle variabili costanti
  null.var = which(apply(X[dati$date == game[match_id], ], 2, sum) ==
    0)
  constant.variable = which(colnames(dati) %in% names(null.var))
  nba.formula = as.formula(paste("y ~ -1 +", paste(colnames(dati[, -c
    (drop.column1, constant.variable, 70)]), collapse = "+")))
  x.game = dati[dati$date == game[match_id], -c(constant.variable, 70)
    ]
  # Utilizzo rstanarm per stimare il modello logistico
  if(match_id == 1) {
    mcmc.rstan = stan_glm(nba.formula, family = binomial(),
      data = dati[dati$date == game[match_id], ], iter =
        niter, warmup = warmup, cores = 4, verbose = F)
  }
  else {
    mcmc.rstan = stan_glm(nba.formula, family = binomial(),
      data = dati[dati$date == game[match_id], ], prior =
        normal(location = coefficients.giornata[match_id-1, -
          null.var], scale = sd.giornata[match_id-1, -null.var]),
      iter = niter, warmup = warmup, cores = 4, verbose = F
    )
  }
  # Aggiornamento delle distribuzioni a priori da fornire al modello
  successivo
  idx = which(colnames(X) %in% names(coef(mcmc.rstan)))
  c.giorn = rep(0, p)
  c.giorn[idx] = mcmc.rstan$coefficients
}
```

```

sd.giorn = rep(2.5,p)
sd.giorn[idx] = se(mcmc.rstan)
if(match_id == 1) {
  coefficienti.giornata[match_id,] = c.giorn
  sd.giornata[match_id,] = sd.giorn
}
else {
  coefficienti.giornata[match_id,] = (1-lambda)*coefficienti
    .giornata[match_id-1,] + lambda*c.giorn
  sd.giornata[match_id,] = (1-lambda)*sd.giornata[match_id
    -1,]+lambda*sd.giorn
}
# Previsioni per i giocatori con i coefficienti stimati
p.lin.pred = posterior_linpred(mcmc.rstan)
pi.chain.action = plogis(p.lin.pred)
p.game.player = apply(pi.chain.action,2,mean)
prev.totali = data.frame(player = as.vector(dati[dati$date == game
  [match_id],"player"]), percentuale = p.game.player)
prev.totali = prev.totali %>%
group_by(player) %>%
summarise(RatingMatch = mean(percentuale),.groups = 'drop')
r.m = data.frame(player = unique(dati$player))
r.m = r.m %>%
left_join(prev.totali,by = c("player" = "player"))
r.m$RatingMatch[is.na(r.m$RatingMatch)] = 0
rank.match[,match_id] = r.m$RatingMatch
cat(match_id," ")
}

```

Codice A.14: Stima del modello logistico con metodi bayesiani dopo ogni giornata, utilizzando degli effetti casuali annidati: un effetto casuale di squadra e un effetto casuale di giocatore, annidato dentro a quello della squadra

```

niter = 4000
warmup = 2000

```

```
game = unique(dati$date)
player.game = unique(dati$player)
num.match = length(game)
num.giocatori = length(player.game)
rank.match = matrix(rep(0,length(game)*length(player.game)), nrow = length
  (player.game), ncol = length(game))
row.names(rank.match) = NULL
colnames(rank.match) = 1:num.match
# Distribuzione a priori utilizzata in questo caso
t_prior = student_t(df = 7, location = 0, scale = 2.5)
for(match_id in 1:num.match) {
  # Rimozione delle variabili costanti
  null.var = which(apply(X[dati$date == game[match_id],],2,sum) ==
    0)
  constant.variable = which(colnames(dati) %in% names(null.var))
  nba.formula = as.formula(paste("y ~ -1 + (1 | team) + (team |
    player) +", paste(colnames(dati[, -c(drop.column, constant.
    variable, 70)]), collapse = "+")))
  x.game = dati[dati$date == game[match_id], -c(constant.variable, 70)
  ]
  # Utilizzo rstanarm per stimare il modello logistico con gli
  effetti casuali annidati
  mcmc.rstan = stan_glmmer(nba.formula, data = dati[dati$date == game
    [match_id],], family = binomial(link = "logit"), prior = t_
    prior, chains = 4, cores = 4, iter = niter)
  # Previsioni per i giocatori
  p.lin.pred = posterior_linpred(mcmc.rstan)
  pi.chain.action = plogis(p.lin.pred)
  p.game.player = apply(pi.chain.action, 2, mean)
  prev.totali = data.frame(player = as.vector(dati[dati$date == game
    [match_id], "player"]), percentuale = p.game.player)
  prev.totali = prev.totali %>%
  group_by(player) %>%
  summarise(RatingMatch = mean(percentuale), .groups = 'drop')
  r.m = data.frame(player = unique(dati$player))
  r.m = r.m %>%
```

```
left_join(prev.totali,by = c("player" = "player"))
r.m$RatingMatch[is.na(r.m$RatingMatch)] = 0
rank.match[,match_id] = r.m$RatingMatch
cat(match_id," ")
}
```

Appendice B

Tabelle dettagliate

In questa Appendice della tesi vengono ora riportate le tabelle che per questioni di spazio non sono state incluse nel corpo principale. In particolare verranno quindi inserite:

- La Tabella [B.1](#) contenente gli indici riassuntivi di base che descrivono le prestazioni dei giocatori presentati nel Capitolo 1;
- La Tabella [B.2](#) contenente gli effetti fissi dei giocatori stimati dal modello di regressione logistica presentato nel Capitolo 2. Questa viene inserita poiché consente di capire e di studiare quanto ciascun giocatore, a prescindere dall'azione che effettua, possa essere considerato pericoloso dal punto di vista offensivo;
- La Tabella [B.3](#) relativa alla classifica di tutti i giocatori NBA selezionati dal modello logistico stimato con il metodo di riduzione della distorsione di Firth presentato nel capitolo 2;
- La Tabella [B.4](#) contenente la classifica dei giocatori NBA che è stata ottenuta dal modello logistico stimato con i metodi *bayesiani* presentati nel Capitolo 3;
- Le Tabelle [B.5](#), [B.6](#), [B.7](#), [B.8](#) e [B.9](#) contenenti le classifiche dei giocatori stratificate per il ruolo di appartenenza, ottenute dall'applicazione di un modello gerarchico stimato con tecniche inferenziali *bayesiane*.

Tabelle dettagliate

Tabella B.1: Indicatori riassuntivi delle performance dei primi 6 giocatori presenti nel dataset

Giocatore	AS	FF	FS	Min	PG	PP	PR	PUN	RD
Aaron Gordon	307	193	285	302.45	83	175	63	1322	469
Aaron Holiday	87	74	57	69.32	52	41	20	299	62
Abdel Nader	20	68	35	55.80	60	26	20	243	102
Al-Farouq Aminu	104	150	173	177.63	81	72	68	760	498
Al Horford	282	132	140	211.63	68	102	58	927	338
Alec Burks	128	92	122	140.32	64	65	39	561	205
RO	RT	SD	SS	T2	T2done	T2miss	T3	T3done	T3miss
140	609	57	54	741	370	371	367	129	238
5	67	13	12	138	63	75	129	44	85
14	116	13	8	117	60	57	100	32	68
112	610	33	41	262	161	101	331	96	235
120	458	87	34	476	314	162	247	73	174
31	236	21	37	306	131	175	168	61	107
TL	TLdone	TLmiss							
272	195	77							
50	41	9							
36	27	9							
173	150	23							
97	80	17							
141	116	25							

Tabella B.2: Stime relative agli effetti fissi di tutti i giocatori NBA ottenuti con il modello di regressione logistica

Parametro	Stima	Standard Error	z value	Pr(> z)
Aaron Gordon	-1.24	0.04	-29.05	0.00
Aaron Holiday	-1.10	0.09	-12.17	0.00
Abdel Nader	-1.41	0.10	-14.26	0.00
Al-Farouq Aminu	-1.15	0.05	-24.16	0.00
Al Horford	-1.19	0.05	-25.87	0.00
Alec Burks	-1.40	0.07	-21.47	0.00
Alex Abrines	-1.48	0.13	-11.63	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Alex Caruso	-1.09	0.10	-10.94	0.00
Alex Len	-1.26	0.05	-23.91	0.00
Alex Poythress	-1.42	0.13	-10.52	0.00
Alfonzo McKinnie	-1.28	0.07	-17.85	0.00
Allen Crabbe	-1.18	0.08	-14.85	0.00
Allonzo Trier	-1.19	0.06	-19.25	0.00
Amir Johnson	-1.36	0.09	-14.86	0.00
Andre Drummond	-1.27	0.04	-33.79	0.00
Andre Iguodala	-1.23	0.05	-22.53	0.00
Andrew Bogut	-1.23	0.12	-10.70	0.00
Andrew Harrison	-1.44	0.18	-7.78	0.00
Andrew Wiggins	-1.23	0.05	-26.19	0.00
Anfernee Simons	-1.61	0.21	-7.80	0.00
Ante Zizic	-1.17	0.07	-17.45	0.00
Anthony Davis	-1.24	0.04	-29.24	0.00
Anthony Tolliver	-1.25	0.08	-14.92	0.00
Antonio Blakeney	-1.26	0.08	-14.99	0.00
Aron Baynes	-1.16	0.07	-16.03	0.00
Austin Rivers	-1.26	0.06	-21.03	0.00
Avery Bradley	-1.27	0.06	-19.85	0.00
Bam Adebayo	-1.14	0.05	-23.09	0.00
Ben McLemore	-0.99	0.20	-5.01	0.00
Ben Simmons	-1.24	0.04	-33.53	0.00
Bismack Biyombo	-1.29	0.08	-16.07	0.00
Blake Griffin	-1.14	0.04	-29.02	0.00
Boban Marjanovic	-1.29	0.07	-19.17	0.00
Bobby Portis	-1.33	0.06	-22.25	0.00
Bogdan Bogdanovic	-1.20	0.05	-22.80	0.00
Bojan Bogdanovic	-1.30	0.05	-27.92	0.00
Bonzie Colson	-0.79	0.22	-3.65	0.00
Brad Wanamaker	-1.15	0.13	-9.17	0.00

Continua nella prossima pagina

Tabella dettagliata

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Bradley Beal	-1.30	0.04	-33.06	0.00
Brandon Ingram	-1.19	0.05	-22.05	0.00
Brandon Knight	-1.33	0.10	-13.87	0.00
Brandon Sampson	-1.42	0.20	-7.01	0.00
Brook Lopez	-1.21	0.05	-26.43	0.00
Bruce Brown	-1.30	0.07	-18.17	0.00
Bruno Caboclo	-1.34	0.09	-15.24	0.00
Bryn Forbes	-1.14	0.05	-21.71	0.00
Buddy Hield	-1.25	0.05	-27.59	0.00
Caleb Swanigan	-1.42	0.17	-8.21	0.00
Cameron Payne	-1.30	0.09	-13.97	0.00
Cameron Reynolds	-1.29	0.17	-7.58	0.00
Caris LeVert	-1.21	0.06	-19.33	0.00
Carmelo Anthony	-1.19	0.14	-8.34	0.00
Cedi Osman	-1.35	0.05	-25.84	0.00
Chandler Hutchison	-1.45	0.09	-15.57	0.00
Chandler Parsons	-1.26	0.11	-11.68	0.00
Channing Frye	-1.48	0.14	-10.34	0.00
Chasson Randle	-1.30	0.09	-13.91	0.00
Cheick Diallo	-1.39	0.07	-18.83	0.00
Chimezie Metu	-1.11	0.18	-6.23	0.00
Chris Boucher	-1.25	0.15	-8.11	0.00
Chris Paul	-1.20	0.04	-27.03	0.00
Christian Wood	-1.52	0.13	-11.83	0.00
CJ McCollum	-1.28	0.04	-29.95	0.00
CJ Miles	-1.20	0.09	-13.69	0.00
Clint Capela	-1.16	0.04	-27.70	0.00
Cody Zeller	-1.20	0.06	-19.47	0.00
Collin Sexton	-1.34	0.05	-27.69	0.00
Corey Brewer	-1.11	0.11	-10.20	0.00
Cory Joseph	-1.20	0.06	-21.51	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Courtney Lee	-1.28	0.12	-10.27	0.00
Cristiano Felicio	-1.29	0.09	-14.74	0.00
D'Angelo Russell	-1.29	0.04	-32.26	0.00
D.J. Augustin	-1.15	0.05	-23.09	0.00
D.J. Wilson	-1.25	0.08	-15.81	0.00
Damian Jones	-1.11	0.11	-9.82	0.00
Damian Lillard	-1.23	0.04	-33.96	0.00
Damion Lee	-1.18	0.13	-9.02	0.00
Damyean Dotson	-1.24	0.06	-20.88	0.00
Daniel Hamilton	-1.31	0.17	-7.83	0.00
Daniel Theis	-1.17	0.07	-16.42	0.00
Danilo Gallinari	-1.19	0.05	-26.14	0.00
Danny Green	-1.12	0.05	-22.19	0.00
Dante Cunningham	-1.16	0.09	-13.11	0.00
Dante Exum	-1.11	0.09	-12.78	0.00
Danuel House Jr.	-1.17	0.08	-14.49	0.00
Dario Saric	-1.24	0.05	-23.29	0.00
Darius Miller	-1.16	0.07	-16.97	0.00
Darren Collison	-1.16	0.05	-23.59	0.00
David Nwaba	-1.17	0.08	-14.65	0.00
Davis Bertans	-1.16	0.06	-18.26	0.00
De'Aaron Fox	-1.14	0.04	-27.45	0.00
De'Anthony Melton	-1.18	0.08	-15.45	0.00
DeAndre' Bembry	-1.28	0.05	-23.93	0.00
Deandre Ayton	-1.29	0.05	-28.29	0.00
DeAndre Jordan	-1.24	0.05	-26.63	0.00
Delon Wright	-1.15	0.06	-20.59	0.00
DeMar DeRozan	-1.27	0.04	-32.70	0.00
DeMarcus Cousins	-1.16	0.06	-19.30	0.00
DeMarre Carroll	-1.23	0.06	-21.63	0.00
Deng Adel	-1.92	0.27	-7.00	0.00

Continua nella prossima pagina

Tabelle dettagliate

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Dennis Schroder	-1.20	0.05	-25.89	0.00
Dennis Smith Jr.	-1.24	0.06	-21.57	0.00
Deonte Burton	-1.17	0.16	-7.22	0.00
Derrick Favors	-1.20	0.05	-25.21	0.00
Derrick Jones Jr.	-1.19	0.07	-16.84	0.00
Derrick Rose	-1.28	0.06	-21.66	0.00
Derrick White	-1.20	0.05	-22.60	0.00
Devin Booker	-1.37	0.04	-31.51	0.00
Devin Harris	-1.19	0.07	-16.38	0.00
Devonte' Graham	-1.13	0.09	-12.06	0.00
Dewayne Dedmon	-1.30	0.05	-24.00	0.00
Deyonta Davis	-1.33	0.22	-6.10	0.00
Dillon Brooks	-1.21	0.14	-8.87	0.00
Dion Waiters	-1.24	0.07	-16.73	0.00
Dirk Nowitzki	-1.32	0.08	-16.11	0.00
Domantas Sabonis	-1.23	0.05	-27.11	0.00
Donovan Mitchell	-1.17	0.04	-29.35	0.00
Donte DiVincenzo	-1.27	0.13	-10.09	0.00
Dorian Finney-Smith	-1.24	0.06	-21.46	0.00
Doug McDermott	-1.15	0.07	-15.78	0.00
Dragan Bender	-1.27	0.09	-14.72	0.00
Draymond Green	-1.21	0.04	-28.81	0.00
Drew Eubanks	-1.34	0.22	-6.11	0.00
Duncan Robinson	-0.89	0.22	-4.11	0.00
Dwayne Bacon	-1.29	0.10	-13.39	0.00
Dwight Howard	-1.17	0.13	-8.75	0.00
Dwight Powell	-1.31	0.05	-23.98	0.00
Dwyane Wade	-1.19	0.05	-24.11	0.00
E'Twaun Moore	-1.27	0.07	-18.26	0.00
Ed Davis	-1.21	0.05	-22.17	0.00
Edmond Sumner	-1.09	0.17	-6.46	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Ekpe Udoh	-1.25	0.12	-10.15	0.00
Elfrid Payton	-1.19	0.06	-19.24	0.00
Elie Okobo	-1.26	0.08	-15.46	0.00
Emmanuel Mudiay	-1.32	0.06	-23.38	0.00
Enes Kanter	-1.41	0.05	-30.76	0.00
Eric Bledsoe	-1.13	0.04	-27.63	0.00
Eric Gordon	-1.28	0.05	-24.71	0.00
Ersan Ilyasova	-1.17	0.06	-19.62	0.00
Evan Fournier	-1.28	0.05	-27.31	0.00
Evan Turner	-1.21	0.05	-22.15	0.00
Frank Jackson	-1.30	0.08	-16.97	0.00
Frank Kaminsky	-1.16	0.08	-14.73	0.00
Frank Mason	-1.34	0.11	-12.39	0.00
Frank Ntilikina	-1.28	0.08	-15.08	0.00
Fred VanVleet	-1.18	0.05	-22.66	0.00
Furkan Korkmaz	-1.21	0.09	-13.45	0.00
Garrett Temple	-1.26	0.06	-20.52	0.00
Gary Clark	-1.32	0.11	-11.94	0.00
Gary Harris	-1.15	0.05	-20.98	0.00
George Hill	-1.15	0.06	-18.98	0.00
Georges Niang	-1.22	0.10	-12.15	0.00
Gerald Green	-1.24	0.07	-18.91	0.00
Giannis Antetokounmpo	-1.05	0.03	-32.43	0.00
Glenn Robinson III	-1.29	0.11	-12.16	0.00
Goran Dragic	-1.08	0.07	-15.47	0.00
Gordon Hayward	-1.22	0.05	-23.99	0.00
Gorgui Dieng	-1.33	0.06	-20.44	0.00
Grayson Allen	-1.55	0.12	-12.80	0.00
Greg Monroe	-1.23	0.08	-14.58	0.00
Guerschon Yabusele	-1.10	0.14	-7.77	0.00
Hamidou Diallo	-1.22	0.10	-12.00	0.00

Continua nella prossima pagina

Tabelle dettagliate

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Harrison Barnes	-1.13	0.05	-22.76	0.00
Harry Giles III	-1.19	0.07	-16.85	0.00
Hassan Whiteside	-1.18	0.05	-25.54	0.00
Henry Ellenson	-1.15	0.14	-8.12	0.00
Ian Clark	-1.19	0.08	-14.61	0.00
Ian Mahinmi	-1.35	0.11	-12.86	0.00
Iman Shumpert	-1.28	0.07	-18.84	0.00
Isaac Bonga	-1.47	0.21	-6.87	0.00
Isaiah Briscoe	-1.04	0.11	-9.70	0.00
Isaiah Canaan	-1.36	0.11	-12.65	0.00
Isaiah Hartenstein	-1.54	0.16	-9.39	0.00
Isaiah Thomas	-1.39	0.16	-8.48	0.00
Ish Smith	-1.15	0.07	-17.61	0.00
Ivan Rabb	-1.27	0.08	-15.40	0.00
Ivica Zubac	-1.32	0.06	-21.21	0.00
J.J. Barea	-1.26	0.07	-17.06	0.00
Jabari Parker	-1.34	0.05	-25.34	0.00
Jacob Evans	-1.37	0.20	-6.92	0.00
Jae Crowder	-1.23	0.05	-24.18	0.00
Jahlil Okafor	-1.31	0.07	-18.66	0.00
JaKarr Sampson	-1.49	0.20	-7.59	0.00
Jake Layman	-1.23	0.07	-17.82	0.00
Jakob Poeltl	-1.33	0.06	-22.32	0.00
Jalen Brunson	-1.37	0.06	-22.03	0.00
Jalen Jones	-1.39	0.17	-7.99	0.00
Jamal Crawford	-1.17	0.07	-17.19	0.00
Jamal Murray	-1.29	0.04	-30.98	0.00
James Ennis III	-1.17	0.07	-17.30	0.00
James Harden	-1.26	0.03	-38.53	0.00
James Johnson	-1.27	0.07	-17.76	0.00
JaMychal Green	-1.26	0.06	-21.73	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Jared Dudley	-1.38	0.08	-17.14	0.00
Jared Terrell	-1.59	0.27	-5.99	0.00
Jarell Martin	-1.33	0.14	-9.73	0.00
Jaren Jackson Jr.	-1.31	0.06	-23.20	0.00
Jaron Blossomgame	-1.48	0.13	-11.13	0.00
Jarrett Allen	-1.30	0.05	-27.75	0.00
Jason Smith	-1.35	0.17	-8.14	0.00
JaVale McGee	-1.28	0.05	-25.72	0.00
Jaylen Adams	-1.52	0.13	-11.90	0.00
Jaylen Brown	-1.14	0.05	-22.80	0.00
Jayson Tatum	-1.28	0.04	-28.89	0.00
Jeff Green	-1.22	0.05	-22.25	0.00
Jeff Teague	-1.23	0.06	-19.69	0.00
Jemerrio Jones	-1.01	0.19	-5.21	0.00
Jerami Grant	-1.17	0.05	-23.95	0.00
Jeremy Lamb	-1.18	0.05	-24.84	0.00
Jeremy Lin	-1.15	0.06	-19.93	0.00
Jerian Grant	-1.26	0.08	-15.07	0.00
Jerome Robinson	-1.02	0.13	-7.79	0.00
Jerryd Bayless	-1.41	0.10	-14.39	0.00
Jevon Carter	-1.37	0.10	-13.19	0.00
Jimmy Butler	-1.30	0.04	-30.18	0.00
JJ Redick	-1.24	0.05	-26.27	0.00
Joakim Noah	-1.14	0.07	-15.26	0.00
Jodie Meeks	-1.59	0.21	-7.41	0.00
Joe Harris	-1.27	0.05	-24.08	0.00
Joe Ingles	-1.16	0.05	-25.54	0.00
Joel Embiid	-1.15	0.04	-32.50	0.00
John Collins	-1.23	0.05	-26.16	0.00
John Henson	-1.17	0.15	-7.95	0.00
John Jenkins	-1.51	0.16	-9.58	0.00

Continua nella prossima pagina

Tabelle dettagiate

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
John Wall	-1.30	0.06	-21.11	0.00
Johnathan Motley	-1.23	0.15	-8.17	0.00
Johnathan Williams	-1.13	0.12	-9.85	0.00
Jon Leuer	-1.38	0.12	-11.48	0.00
Jonah Bolden	-1.24	0.09	-14.10	0.00
Jonas Jerebko	-1.14	0.06	-17.69	0.00
Jonas Valanciunas	-1.33	0.06	-23.27	0.00
Jonathan Isaac	-1.27	0.05	-23.74	0.00
Jonathon Simmons	-1.33	0.07	-18.09	0.00
Jordan Bell	-1.20	0.08	-15.40	0.00
Jordan Clarkson	-1.27	0.05	-25.61	0.00
Jordan McRae	-1.11	0.13	-8.55	0.00
Jose Calderon	-1.25	0.11	-11.62	0.00
Josh Hart	-1.20	0.07	-18.46	0.00
Josh Jackson	-1.27	0.05	-24.92	0.00
Josh Okogie	-1.22	0.06	-19.88	0.00
Josh Richardson	-1.16	0.05	-24.43	0.00
JR Smith	-1.02	0.17	-6.09	0.00
Jrue Holiday	-1.13	0.04	-26.36	0.00
Juancho Hernangomez	-1.06	0.07	-14.71	0.00
Julian Washburn	-1.52	0.19	-8.21	0.00
Julius Randle	-1.24	0.04	-29.66	0.00
Justin Anderson	-1.24	0.11	-11.39	0.00
Justin Holiday	-1.23	0.05	-22.88	0.00
Justin Jackson	-1.25	0.07	-18.36	0.00
Justise Winslow	-1.17	0.05	-22.74	0.00
Jusuf Nurkic	-1.19	0.04	-28.05	0.00
Kadeem Allen	-1.19	0.11	-10.97	0.00
Karl-Anthony Towns	-1.17	0.04	-31.32	0.00
Kawhi Leonard	-1.22	0.04	-32.62	0.00
Keita Bates-Diop	-1.37	0.12	-11.38	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Kelly Olynyk	-1.28	0.06	-23.04	0.00
Kelly Oubre Jr.	-1.10	0.05	-21.13	0.00
Kemba Walker	-1.16	0.04	-29.56	0.00
Kenneth Faried	-1.21	0.07	-16.22	0.00
Kenrich Williams	-1.35	0.08	-16.79	0.00
Kent Bazemore	-1.31	0.06	-23.50	0.00
Kentavious Caldwell-Pope	-1.26	0.06	-21.97	0.00
Kevin Durant	-1.20	0.04	-32.53	0.00
Kevin Huerter	-1.17	0.06	-20.83	0.00
Kevin Knox	-1.30	0.05	-24.27	0.00
Kevin Love	-1.36	0.08	-16.90	0.00
Kevon Looney	-1.18	0.05	-22.19	0.00
Khem Birch	-1.23	0.08	-14.83	0.00
Khris Middleton	-1.09	0.04	-27.17	0.00
Khyri Thomas	-1.41	0.19	-7.47	0.00
Klay Thompson	-1.17	0.04	-28.83	0.00
Kosta Koufos	-1.30	0.10	-13.48	0.00
Kris Dunn	-1.28	0.06	-20.97	0.00
Kyle Anderson	-1.38	0.07	-19.46	0.00
Kyle Korver	-1.16	0.07	-17.04	0.00
Kyle Kuzma	-1.12	0.05	-23.43	0.00
Kyle Lowry	-1.08	0.04	-26.55	0.00
Kyle O'Quinn	-1.02	0.10	-9.94	0.00
Kyrie Irving	-1.17	0.04	-28.71	0.00
LaMarcus Aldridge	-1.30	0.04	-33.53	0.00
Lance Stephenson	-1.27	0.07	-18.64	0.00
Lance Thomas	-1.49	0.10	-14.62	0.00
Landry Shamet	-1.20	0.06	-19.34	0.00
Langston Galloway	-1.21	0.06	-18.68	0.00
Larry Nance Jr.	-1.39	0.05	-26.27	0.00
Lauri Markkanen	-1.24	0.05	-23.27	0.00

Continua nella prossima pagina

Tabella dettagliata

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
LeBron James	-1.13	0.04	-26.63	0.00
Lonnie Walker IV	-1.46	0.21	-6.79	0.00
Lonzo Ball	-1.22	0.06	-19.67	0.00
Lorenzo Brown	-1.02	0.16	-6.30	0.00
Lou Williams	-1.28	0.04	-29.99	0.00
Luka Doncic	-1.30	0.04	-31.68	0.00
Luke Kennard	-1.19	0.07	-18.11	0.00
Luke Kornet	-1.26	0.09	-14.59	0.00
Luol Deng	-1.20	0.12	-9.63	0.00
Malcolm Brogdon	-1.16	0.05	-22.98	0.00
Malik Beasley	-1.20	0.06	-21.18	0.00
Malik Monk	-1.21	0.07	-18.36	0.00
Marc Gasol	-1.20	0.04	-30.26	0.00
Marcin Gortat	-1.52	0.08	-18.33	0.00
Marco Belinelli	-1.20	0.06	-20.50	0.00
Marcus Morris	-1.26	0.05	-25.94	0.00
Marcus Smart	-1.17	0.05	-22.38	0.00
Mario Hezonja	-1.08	0.07	-16.61	0.00
Markelle Fultz	-1.17	0.11	-10.78	0.00
Markieff Morris	-1.19	0.06	-18.59	0.00
Marquese Chriss	-1.27	0.10	-12.84	0.00
MarShon Brooks	-1.39	0.12	-11.25	0.00
Marvin Bagley III	-1.20	0.05	-22.85	0.00
Marvin Williams	-1.20	0.06	-21.76	0.00
Mason Plumlee	-1.14	0.05	-23.93	0.00
Matthew Dellavedova	-1.22	0.08	-14.46	0.00
Maurice Harkless	-1.17	0.06	-20.35	0.00
Maxi Kleber	-1.37	0.07	-20.95	0.00
Meyers Leonard	-1.25	0.07	-17.31	0.00
Michael Beasley	-1.16	0.12	-9.97	0.00
Michael Carter-Williams	-1.19	0.10	-12.23	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Michael Kidd-Gilchrist	-1.13	0.07	-16.81	0.00
Mikal Bridges	-1.28	0.06	-22.44	0.00
Mike Conley	-1.23	0.04	-27.99	0.00
Mike Muscala	-1.12	0.07	-16.27	0.00
Mike Scott	-1.14	0.07	-17.43	0.00
Miles Bridges	-1.24	0.06	-20.38	0.00
Miles Plumlee	-1.31	0.16	-8.02	0.00
Milos Teodosic	-1.32	0.19	-6.88	0.00
Mitchell Robinson	-1.27	0.06	-21.05	0.00
Mo Bamba	-1.43	0.09	-16.36	0.00
Mohamed Bamba	-1.29	0.20	-6.39	0.00
Monte Morris	-1.08	0.05	-20.43	0.00
Montrezl Harrell	-1.27	0.04	-29.21	0.00
Moritz Wagner	-1.32	0.11	-12.28	0.00
Myles Turner	-1.25	0.05	-26.54	0.00
Nemanja Bjelica	-1.19	0.05	-22.05	0.00
Nene	-1.23	0.10	-12.70	0.00
Nerlens Noel	-1.22	0.06	-18.98	0.00
Nicolas Batum	-1.28	0.05	-23.37	0.00
Nik Stauskas	-1.12	0.08	-14.14	0.00
Nikola Jokic	-1.19	0.03	-34.78	0.00
Nikola Mirotic	-1.22	0.06	-21.81	0.00
Nikola Vucevic	-1.20	0.04	-31.61	0.00
Noah Vonleh	-1.35	0.06	-24.31	0.00
Norman Powell	-1.03	0.06	-16.33	0.00
OG Anunoby	-1.11	0.07	-15.60	0.00
Omari Spellman	-1.42	0.09	-16.66	0.00
Omri Casspi	-0.98	0.10	-9.90	0.00
Otto Porter Jr.	-1.24	0.06	-21.17	0.00
Pascal Siakam	-1.16	0.04	-30.12	0.00
Pat Connaughton	-1.21	0.06	-19.24	0.00

Continua nella prossima pagina

Tabelle dettagliate

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Patrick Beverley	-1.15	0.05	-22.53	0.00
Patrick McCaw	-1.23	0.14	-8.99	0.00
Patrick Patterson	-1.28	0.09	-13.68	0.00
Patty Mills	-1.23	0.06	-21.92	0.00
Pau Gasol	-1.05	0.11	-9.97	0.00
Paul George	-1.19	0.04	-32.26	0.00
Paul Millsap	-1.20	0.05	-26.22	0.00
PJ Tucker	-1.18	0.05	-23.10	0.00
Quincy Acy	-1.11	0.23	-4.90	0.00
Quincy Pondexter	-0.87	0.14	-6.06	0.00
Quinn Cook	-1.36	0.07	-19.24	0.00
Rajon Rondo	-1.31	0.06	-21.28	0.00
Raul Neto	-1.10	0.10	-11.03	0.00
Rawle Alkins	-1.17	0.21	-5.50	0.00
Ray Spalding	-1.21	0.17	-7.10	0.00
Raymond Felton	-1.26	0.12	-10.57	0.00
Reggie Bullock	-1.38	0.07	-20.97	0.00
Reggie Jackson	-1.20	0.05	-25.74	0.00
Richaun Holmes	-1.27	0.06	-20.64	0.00
Ricky Rubio	-1.25	0.05	-26.24	0.00
Robert Covington	-1.26	0.07	-17.56	0.00
Robert Williams III	-1.56	0.14	-11.21	0.00
Robin Lopez	-1.25	0.06	-20.96	0.00
Rodions Kurucs	-1.39	0.07	-21.03	0.00
Rodney Hood	-1.30	0.06	-23.35	0.00
Rodney McGruder	-1.15	0.07	-17.34	0.00
Rondae Hollis-Jefferson	-1.20	0.06	-20.24	0.00
Royce O'Neale	-1.21	0.06	-19.69	0.00
Rudy Gay	-1.22	0.05	-25.13	0.00
Rudy Gobert	-1.19	0.04	-30.54	0.00
Russell Westbrook	-1.18	0.03	-34.03	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Ryan Anderson	-1.24	0.15	-8.04	0.00
Ryan Arcidiacono	-1.17	0.06	-19.20	0.00
Ryan Broekhoff	-1.20	0.13	-9.17	0.00
Salah Mejri	-1.32	0.11	-12.36	0.00
Sam Dekker	-1.26	0.09	-14.30	0.00
Semi Ojeleye	-1.07	0.11	-10.07	0.00
Serge Ibaka	-1.20	0.04	-28.05	0.00
Seth Curry	-1.28	0.07	-18.88	0.00
Shabazz Napier	-1.09	0.07	-15.93	0.00
Shai Gilgeous-Alexander	-1.27	0.05	-25.49	0.00
Shake Milton	-1.33	0.16	-8.19	0.00
Shaquille Harrison	-1.37	0.06	-21.14	0.00
Shaun Livingston	-1.23	0.07	-17.24	0.00
Shelvin Mack	-1.27	0.07	-17.90	0.00
Sindarius Thornwell	-1.22	0.14	-8.62	0.00
Skal Labissiere	-1.38	0.18	-7.73	0.00
Solomon Hill	-1.26	0.10	-13.21	0.00
Spencer Dinwiddie	-1.30	0.05	-26.00	0.00
Stanley Johnson	-1.28	0.07	-18.47	0.00
Stephen Curry	-1.21	0.04	-32.25	0.00
Sterling Brown	-1.21	0.07	-16.86	0.00
Steven Adams	-1.18	0.04	-27.54	0.00
Svi Mykhailiuk	-1.20	0.13	-9.17	0.00
T.J. McConnell	-1.02	0.06	-16.68	0.00
T.J. Warren	-1.39	0.07	-20.17	0.00
Taj Gibson	-1.20	0.05	-21.93	0.00
Taurean Prince	-1.27	0.06	-20.57	0.00
Terrance Ferguson	-1.15	0.07	-17.11	0.00
Terrence Ross	-1.24	0.05	-24.39	0.00
Terry Rozier	-1.26	0.05	-23.29	0.00
Thabo Sefolosha	-1.43	0.10	-14.05	0.00

Continua nella prossima pagina

Tabelle dettagliate

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Thaddeus Young	-1.19	0.05	-25.89	0.00
Theo Pinson	-1.30	0.16	-8.02	0.00
Thomas Bryant	-1.27	0.06	-22.32	0.00
Thon Maker	-1.39	0.08	-18.02	0.00
Tim Frazier	-1.17	0.07	-16.63	0.00
Tim Hardaway Jr.	-1.31	0.05	-24.95	0.00
Timothe Luwawu-Cabarrot	-1.24	0.10	-12.24	0.00
TJ Leaf	-1.07	0.10	-10.77	0.00
Tobias Harris	-1.23	0.04	-30.92	0.00
Tomas Satoransky	-1.28	0.05	-24.22	0.00
Tony Parker	-1.17	0.07	-17.25	0.00
Tony Snell	-1.06	0.07	-14.31	0.00
Torrey Craig	-1.25	0.06	-20.22	0.00
Trae Young	-1.27	0.04	-31.39	0.00
Treveon Graham	-1.17	0.10	-12.03	0.00
Trevor Ariza	-1.34	0.05	-25.38	0.00
Trey Burke	-1.18	0.07	-17.60	0.00
Trey Lyles	-1.39	0.07	-20.62	0.00
Tristan Thompson	-1.39	0.06	-21.82	0.00
Troy Brown Jr.	-1.12	0.09	-12.60	0.00
Troy Daniels	-1.23	0.10	-12.37	0.00
Troy Williams	-1.23	0.14	-8.60	0.00
Tyler Dorsey	-1.25	0.09	-13.70	0.00
Tyler Johnson	-1.14	0.06	-18.13	0.00
Tyler Zeller	-1.39	0.22	-6.21	0.00
Tyreke Evans	-1.10	0.06	-19.58	0.00
Tyrone Wallace	-1.11	0.09	-12.04	0.00
Tyson Chandler	-1.25	0.08	-15.66	0.00
Tyus Jones	-1.18	0.06	-18.73	0.00
Victor Oladipo	-1.09	0.06	-17.84	0.00
Vince Carter	-1.36	0.07	-19.63	0.00

Continua nella prossima pagina

Parametro giocatore	Stima	Standard Error	z value	Pr(> z)
Walter Lemon Jr.	-1.50	0.17	-8.98	0.00
Wayne Ellington	-1.28	0.07	-17.43	0.00
Wayne Selden	-1.23	0.07	-18.33	0.00
Wendell Carter Jr.	-1.42	0.07	-21.44	0.00
Wes Iwundu	-1.26	0.07	-16.90	0.00
Wesley Johnson	-1.36	0.12	-11.12	0.00
Wesley Matthews	-1.27	0.06	-22.21	0.00
Will Barton	-1.20	0.06	-19.60	0.00
Willie Cauley-Stein	-1.27	0.05	-27.70	0.00
Willy Hernangomez	-1.15	0.07	-16.66	0.00
Wilson Chandler	-1.39	0.08	-17.93	0.00
Yogi Ferrell	-1.10	0.08	-14.60	0.00
Yuta Watanabe	-0.94	0.19	-4.84	0.00
Zach Collins	-1.09	0.06	-19.33	0.00
Zach LaVine	-1.34	0.05	-29.03	0.00
Zaza Pachulia	-1.30	0.07	-17.96	0.00

Tabella B.3: Classifica di tutti i giocatori di NBA selezionati dal modello logistico stimato con il metodo di riduzione della distorsione di Firth

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
1	Giannis Antetokounmpo	MIL	PF	0.40	23.77
2	James Harden	HOU	PG	0.37	23.00
3	Anthony Davis	NOP	C	0.42	22.07
4	Nikola Jokic	DEN	C	0.40	20.20
5	Russell Westbrook	OKC	PG	0.32	19.59
6	Luka Doncic	DAL	SG	0.41	19.08
7	LeBron James	LAL	SF	0.35	19.08
8	Joel Embiid	PHI	C	0.31	18.50
9	Jrue Holiday	NOP	SG	0.39	17.60

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
10	Paul George	OKC	SF	0.34	17.41
11	Trae Young	ATL	PG	0.40	17.36
12	Stephen Curry	GSW	PG	0.39	17.13
13	Karl-Anthony Towns	MIN	C	0.33	17.08
14	Buddy Hield	SAC	SG	0.49	16.70
15	Kawhi Leonard	TOR	SF	0.35	16.51
16	Bradley Beal	WAS	SG	0.37	16.43
17	Jusuf Nurkic	POR	C	0.39	16.34
18	Kevin Durant	GSW	SF	0.36	16.28
19	DeMar DeRozan	SAS	SG	0.36	16.02
20	Blake Griffin	DET	PF	0.34	15.93
21	Kemba Walker	CHA	PG	0.36	15.78
22	DeMarcus Cousins	GSW	C	0.40	15.67
23	Kevin Love	CLE	PF	0.40	15.45
24	Damian Lillard	POR	PG	0.34	15.27
25	Mike Conley	MEM	PG	0.38	15.24
26	Ben Simmons	PHI	PG	0.34	15.20
27	Devin Booker	PHX	SG	0.30	14.55
28	Nikola Vucevic	ORL	C	0.33	14.42
29	De'Aaron Fox	SAC	PG	0.36	14.33
30	Lonzo Ball	LAL	PG	0.46	14.13
31	Julius Randle	NOP	PF	0.32	14.06
32	Pascal Siakam	TOR	PF	0.38	13.75
33	Donovan Mitchell	UTA	SG	0.31	13.74
34	Zach LaVine	CHI	SG	0.32	13.72
35	Hassan Whiteside	MIA	C	0.39	13.68
36	Kyle Lowry	TOR	PG	0.36	13.65
37	Andre Drummond	DET	C	0.28	13.24
38	Chris Paul	HOU	PG	0.32	13.17
39	Lauri Markkanen	CHI	PF	0.35	13.07
40	Justise Winslow	MIA	SF	0.41	13.06

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
41	Brandon Ingram	LAL	SF	0.35	12.92
42	Kyrie Irving	BOS	PG	0.29	12.71
43	Danilo Gallinari	LAC	SF	0.36	12.68
44	Klay Thompson	GSW	SG	0.36	12.64
45	Kyle Kuzma	LAL	PF	0.36	12.58
46	John Wall	WAS	PG	0.27	12.57
47	Dennis Schroder	OKC	PG	0.39	12.57
48	Serge Ibaka	TOR	C	0.40	12.50
49	Goran Dragic	MIA	PG	0.40	12.27
50	Thaddeus Young	IND	PF	0.40	12.23
51	CJ McCollum	POR	SG	0.33	12.17
52	Andrew Wiggins	MIN	SF	0.34	12.10
53	Deandre Ayton	PHX	C	0.32	12.10
54	John Collins	ATL	PF	0.30	12.09
55	Elfrid Payton	NOP	PG	0.36	12.09
56	Derrick Favors	UTA	C	0.41	12.09
57	Ricky Rubio	UTA	PG	0.35	12.06
58	Victor Oladipo	IND	SG	0.30	12.05
59	Khris Middleton	MIL	SF	0.32	11.98
60	Jamal Murray	DEN	PG	0.32	11.86
61	Clint Capela	HOU	C	0.31	11.81
62	Draymond Green	GSW	PF	0.32	11.72
63	Tristan Thompson	CLE	C	0.37	11.66
64	Taurean Prince	ATL	SF	0.42	11.59
65	Jeremy Lamb	CHA	SG	0.38	11.51
66	Josh Richardson	MIA	SG	0.34	11.49
67	Willie Cauley-Stein	SAC	C	0.35	11.44
68	Al Horford	BOS	C	0.34	11.35
69	Rudy Gobert	UTA	C	0.27	11.20
70	Domantas Sabonis	IND	C	0.32	10.96
71	Jeff Teague	MIN	PG	0.33	10.96

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
72	Jarrett Allen	BKN	C	0.37	10.94
73	Dwyane Wade	MIA	SG	0.34	10.93
74	Jaylen Brown	BOS	SG	0.40	10.93
75	Walter Lemon Jr.	CHI	PG	0.32	10.93
76	Paul Millsap	DEN	PF	0.35	10.88
77	Lou Williams	LAC	SG	0.29	10.82
78	Joe Ingles	UTA	PF	0.35	10.80
79	Rajon Rondo	LAL	PG	0.34	10.76
80	Caris LeVert	BKN	SF	0.35	10.74
81	Tomas Satoransky	WAS	PG	0.43	10.54
82	LaMarcus Aldridge	SAS	C	0.25	10.52
83	Steven Adams	OKC	C	0.31	10.52
84	JaKarr Sampson	CHI	SF	0.30	10.42
85	Marcus Morris	BOS	PF	0.36	10.23
86	Rudy Gay	SAS	PF	0.33	10.15
87	Eric Bledsoe	MIL	PG	0.29	10.13
88	Malcolm Brogdon	MIL	SG	0.34	10.00
89	Jerami Grant	OKC	PF	0.37	9.94
90	Derrick Rose	MIN	PG	0.33	9.93
91	Collin Sexton	CLE	PG	0.32	9.87
92	JaVale McGee	LAL	C	0.32	9.78
93	Marvin Bagley III	SAC	PF	0.31	9.65
94	D'Angelo Russell	BKN	PG	0.23	9.57
95	Aaron Gordon	ORL	PF	0.26	9.54
96	Jayson Tatum	BOS	SF	0.29	9.42
97	Kris Dunn	CHI	PG	0.28	9.34
98	Josh Jackson	PHX	SG	0.32	9.25
99	Bojan Bogdanovic	IND	SF	0.30	9.24
100	Kadeem Allen	NYK	SG	0.38	9.21
101	Shabazz Napier	BKN	PG	0.47	9.19
102	Jeff Green	WAS	PF	0.38	9.18

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
103	Gary Harris	DEN	SG	0.35	9.10
104	Larry Nance Jr.	CLE	C	0.31	9.07
105	Tony Parker	CHA	PG	0.44	9.05
106	Emmanuel Mudiay	NYK	PG	0.29	9.05
107	Jordan Clarkson	CLE	SG	0.32	9.05
108	Fred VanVleet	TOR	PG	0.39	9.03
109	J.J. Barea	DAL	PG	0.34	9.02
110	Brook Lopez	MIL	C	0.31	8.98
111	Darren Collison	IND	PG	0.32	8.97
112	Alex Len	ATL	C	0.34	8.96
113	Jae Crowder	UTA	SF	0.34	8.92
114	Jaren Jackson Jr.	MEM	PF	0.28	8.82
115	Reggie Jackson	DET	PG	0.29	8.82
116	Patrick Beverley	LAC	PG	0.34	8.81
117	Thomas Bryant	WAS	C	0.38	8.75
118	D.J. Augustin	ORL	PG	0.35	8.72
119	Nicolas Batum	CHA	SF	0.35	8.66
120	Dwight Powell	DAL	C	0.36	8.64
121	Montrezl Harrell	LAC	C	0.26	8.63
122	Rodions Kurucs	BKN	SF	0.42	8.59
123	Bogdan Bogdanovic	SAC	SG	0.29	8.52
124	Jemerrio Jones	LAL	SF	0.39	8.47
125	Mikal Bridges	PHX	SF	0.39	8.45
126	Kyle Anderson	MEM	SF	0.32	8.36
127	Marcus Smart	BOS	SG	0.33	8.33
128	Derrick White	SAS	PG	0.31	8.32
129	Bam Adebayo	MIA	C	0.31	8.30
130	OG Anunoby	TOR	SF	0.50	8.24
131	PJ Tucker	HOU	PF	0.35	8.15
132	Eric Gordon	HOU	SG	0.29	8.14
133	Kevin Knox	NYK	PF	0.29	8.07

Continua nella prossima pagina

Tabella dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
134	Al-Farouq Aminu	POR	PF	0.33	8.03
135	Rondae Hollis-Jefferson	BKN	SF	0.33	8.00
136	Darius Miller	NOP	SF	0.44	7.92
137	Marc Gasol	TOR	C	0.45	7.78
138	Taj Gibson	MIN	PF	0.30	7.78
139	Gordon Hayward	BOS	PF	0.30	7.74
140	Danny Green	TOR	SG	0.37	7.72
141	JJ Redick	PHI	SG	0.26	7.70
142	Allonzo Trier	NYK	SG	0.34	7.69
143	Tobias Harris	PHI	PF	0.41	7.69
144	Cody Zeller	CHA	C	0.26	7.69
145	Nemanja Bjelica	SAC	PF	0.31	7.66
146	Myles Turner	IND	C	0.23	7.65
147	E'Twaun Moore	NOP	SG	0.34	7.65
148	Jimmy Butler	MIN	SF	0.40	7.64
149	Johnathan Williams	LAL	C	0.44	7.57
150	Joakim Noah	MEM	C	0.33	7.51
151	Andre Iguodala	GSW	SF	0.37	7.50
152	Tyreke Evans	IND	SG	0.31	7.49
153	Alex Caruso	LAL	PG	0.32	7.49
154	James Johnson	MIA	PF	0.37	7.47
155	Miles Bridges	CHA	SF	0.40	7.45
156	Noah Vonleh	NYK	PF	0.28	7.44
157	Spencer Dinwiddie	BKN	PG	0.24	7.44
158	Derrick Jones Jr.	MIA	SF	0.40	7.39
159	Mason Plumlee	DEN	C	0.29	7.38
160	Dwight Howard	WAS	C	0.22	7.36
161	Kentavious Caldwell-Pope	LAL	SG	0.35	7.35
162	Kent Bazemore	ATL	SG	0.26	7.35
163	Shai Gilgeous-Alexander	LAC	PG	0.28	7.27
164	De'Anthony Melton	PHX	PG	0.37	7.24

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
165	Jahlil Okafor	NOP	C	0.37	7.24
166	Bruno Caboclo	MEM	SF	0.33	7.19
167	Maurice Harkless	POR	SF	0.32	7.12
168	Will Barton	DEN	SF	0.27	7.11
169	Dorian Finney-Smith	DAL	SF	0.34	7.09
170	Terrence Ross	ORL	SG	0.28	7.06
171	Tyus Jones	MIN	PG	0.36	7.06
172	Enes Kanter	POR	C	0.43	7.03
173	Mitchell Robinson	NYK	C	0.30	7.01
174	Ryan Arcidiacono	CHI	PG	0.38	6.96
175	Allen Crabbe	BKN	SG	0.32	6.92
176	Evan Fournier	ORL	SG	0.22	6.91
177	Bryn Forbes	SAS	SG	0.30	6.89
178	Treveon Graham	BKN	SG	0.45	6.88
179	Monte Morris	DEN	PG	0.35	6.87
180	Josh Okogie	MIN	SG	0.32	6.86
181	Josh Hart	LAL	SG	0.34	6.80
182	Richaun Holmes	PHX	C	0.32	6.80
183	Evan Turner	POR	PG	0.33	6.77
184	Jamal Crawford	PHX	SG	0.36	6.73
185	Markelle Fultz	PHI	SG	0.27	6.72
186	Kelly Olynyk	MIA	PF	0.29	6.66
187	Dion Waiters	MIA	SG	0.28	6.65
188	Kevin Huerter	ATL	SG	0.28	6.62
189	Dragan Bender	PHX	PF	0.38	6.57
190	Wes Iwundu	ORL	SF	0.47	6.52
191	Omri Casspi	MEM	SF	0.46	6.51
192	Kenrich Williams	NOP	SF	0.33	6.51
193	Marco Belinelli	SAS	SG	0.33	6.39
194	Patty Mills	SAS	PG	0.31	6.32
195	Jalen Brunson	DAL	PG	0.30	6.27

Continua nella prossima pagina

Tabella dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
196	Royce O'Neale	UTA	SF	0.36	6.25
197	DeAndre' Bembry	ATL	SG	0.25	6.21
198	Chandler Parsons	MEM	SF	0.31	6.21
199	Torrey Craig	DEN	SF	0.36	6.14
200	Joe Harris	BKN	SG	0.24	6.08
201	Terry Rozier	BOS	PG	0.28	6.06
202	Ersan Ilyasova	MIL	PF	0.31	6.05
203	Maxi Kleber	DAL	PF	0.31	5.94
204	Mohamed Bamba	ORL	C	0.31	5.93
205	Frank Kaminsky	CHA	C	0.31	5.92
206	Wendell Carter Jr.	CHI	C	0.19	5.91
207	Malik Monk	CHA	SG	0.33	5.90
208	Mario Hezonja	NYK	SF	0.27	5.88
209	Frank Ntilikina	NYK	PG	0.30	5.88
210	Devin Harris	DAL	PG	0.37	5.87
211	T.J. Warren	PHX	SF	0.19	5.80
212	Dwayne Bacon	CHA	SG	0.37	5.79
213	Luke Kennard	DET	SG	0.29	5.66
214	Salah Mejri	DAL	C	0.39	5.65
215	Dewayne Dedmon	ATL	C	0.19	5.64
216	Ante Zizic	CLE	C	0.27	5.53
217	Luke Kornet	NYK	PF	0.33	5.51
218	Dario Saric	PHI	PF	0.44	5.42
219	Nerlens Noel	OKC	C	0.33	5.41
220	Antonio Blakeney	CHI	SG	0.37	5.40
221	DeAndre Jordan	NYK	C	0.30	5.38
222	Pat Connaughton	MIL	SG	0.30	5.34
223	Lance Stephenson	LAL	SG	0.29	5.31
224	Dirk Nowitzki	DAL	PF	0.31	5.29
225	Robin Lopez	CHI	C	0.25	5.23
226	Robert Covington	PHI	SF	0.33	5.16

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
227	Trevor Ariza	PHX	SF	0.35	5.16
228	Tim Hardaway Jr.	DAL	SG	0.32	5.15
229	Elie Okobo	PHX	PG	0.30	5.13
230	Isaiah Thomas	DEN	SG	0.27	5.08
231	Cedi Osman	CLE	SF	0.18	5.04
232	Jevon Carter	MEM	PG	0.35	5.01
233	Ian Clark	NOP	SG	0.34	5.01
234	Omari Spellman	ATL	PF	0.29	4.99
235	Jonathan Isaac	ORL	PF	0.20	4.95
236	Cheick Diallo	NOP	PF	0.29	4.93
237	Frank Jackson	NOP	PG	0.30	4.93
238	Kevon Looney	GSW	C	0.26	4.89
239	Carmelo Anthony	HOU	PF	0.17	4.86
240	Gorgui Dieng	MIN	C	0.28	4.86
241	Bobby Portis	CHI	PF	0.30	4.84
242	Marvin Williams	CHA	PF	0.20	4.81
243	Dennis Smith Jr.	NYK	PG	0.29	4.79
244	Cristiano Felicio	CHI	C	0.38	4.68
245	Kelly Oubre Jr.	WAS	SF	0.31	4.64
246	Khem Birch	ORL	C	0.32	4.63
247	Ish Smith	DET	PG	0.21	4.60
248	Damyean Dotson	NYK	SG	0.21	4.60
249	Rawle Alkins	CHI	SG	0.39	4.55
250	Shaquille Harrison	CHI	SG	0.24	4.52
251	Jonas Valanciunas	TOR	C	0.26	4.51
252	Alex Poythress	ATL	PF	0.29	4.50
253	Mo Bamba	ORL	C	0.23	4.47
254	Jared Dudley	BKN	PF	0.30	4.46
255	Jabari Parker	WAS	PF	0.27	4.35
256	Otto Porter Jr.	CHI	SF	0.31	4.34
257	Quinn Cook	GSW	PG	0.33	4.34

Continua nella prossima pagina

Tabella dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
258	Dillon Brooks	MEM	SF	0.24	4.32
259	Delon Wright	TOR	PG	0.38	4.31
260	Meyers Leonard	POR	C	0.29	4.26
261	Trey Burke	DAL	PG	0.42	4.25
262	Ivan Rabb	MEM	PF	0.25	4.23
263	Cory Joseph	IND	PG	0.21	4.22
264	DeMarre Carroll	BKN	PF	0.18	4.22
265	Jordan McRae	WAS	SG	0.32	4.21
266	Bismack Biyombo	CHA	C	0.26	4.16
267	Harrison Barnes	SAC	PF	0.29	4.12
268	Donte DiVincenzo	MIL	SG	0.31	4.11
269	Ed Davis	BKN	C	0.19	4.11
270	Cameron Payne	CLE	PG	0.50	4.09
271	Moritz Wagner	LAL	C	0.33	4.06
272	Yogi Ferrell	SAC	PG	0.31	4.05
273	Devonte' Graham	CHA	PG	0.30	4.04
274	Rodney McGruder	MIA	SG	0.21	3.99
275	JaMychal Green	LAC	PF	0.33	3.96
276	John Henson	MIL	C	0.23	3.95
277	Damion Lee	GSW	SG	0.40	3.94
278	Lance Thomas	NYK	PF	0.30	3.93
279	Avery Bradley	MEM	SG	0.33	3.90
280	Zach Collins	POR	C	0.20	3.86
281	Troy Brown Jr.	WAS	SF	0.29	3.84
282	Michael Beasley	LAL	PF	0.24	3.83
283	Keita Bates-Diop	MIN	SF	0.28	3.82
284	TJ Leaf	IND	PF	0.44	3.81
285	Jakob Poeltl	SAS	C	0.21	3.80
286	Robert Williams III	BOS	C	0.38	3.78
287	Chandler Hutchison	CHI	SF	0.25	3.77
288	Harry Giles III	SAC	PF	0.19	3.71

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
289	Yuta Watanabe	MEM	SF	0.40	3.70
290	Kosta Koufos	SAC	C	0.26	3.70
291	Justin Anderson	ATL	SF	0.34	3.69
292	Ray Spalding	PHX	PF	0.25	3.69
293	Bonzie Colson	MIL	SF	0.25	3.67
294	Jerome Robinson	LAC	SG	0.41	3.66
295	Isaiah Briscoe	ORL	PG	0.29	3.65
296	Jalen Jones	CLE	SF	0.28	3.61
297	Norman Powell	TOR	SG	0.22	3.59
298	Bruce Brown	DET	SG	0.24	3.58
299	Nik Stauskas	CLE	SG	0.57	3.58
300	Davis Bertans	SAS	PF	0.21	3.58
301	David Nwaba	CLE	SG	0.21	3.57
302	Tyler Johnson	PHX	PG	0.30	3.56
303	Chasson Randle	WAS	PG	0.26	3.56
304	Jerian Grant	ORL	PG	0.28	3.55
305	Johnathan Motley	LAC	PF	0.32	3.55
306	Justin Holiday	CHI	SG	0.29	3.53
307	Furkan Korkmaz	PHI	SG	0.26	3.52
308	George Hill	CLE	PG	0.36	3.52
309	Jerryd Bayless	MIN	PG	0.19	3.51
310	Quincy Acy	PHX	PF	0.31	3.51
311	Shaun Livingston	GSW	PG	0.28	3.51
312	Terrance Ferguson	OKC	SG	0.21	3.50
313	Brandon Sampson	CHI	SG	0.30	3.47
314	Seth Curry	POR	SG	0.25	3.45
315	Nene	HOU	C	0.28	3.43
316	Solomon Hill	NOP	SF	0.24	3.42
317	Michael Kidd-Gilchrist	CHA	PF	0.18	3.41
318	Shake Milton	PHI	SG	0.30	3.41
319	Tyler Zeller	ATL	C	0.35	3.39

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
320	JR Smith	CLE	SG	0.19	3.36
321	Tyrone Wallace	LAC	PG	0.30	3.36
322	Miles Plumlee	ATL	C	0.29	3.35
323	Frank Mason	SAC	PG	0.25	3.30
324	Stanley Johnson	NOP	SF	0.35	3.23
325	Georges Niang	UTA	PF	0.36	3.22
326	Gerald Green	HOU	SG	0.20	3.20
327	Ian Mahinmi	WAS	C	0.21	3.19
328	Duncan Robinson	MIA	PF	0.37	3.19
329	Reggie Bullock	LAL	SG	0.29	3.16
330	James Ennis III	PHI	SF	0.34	3.05
331	Luol Deng	MIN	SF	0.20	3.03
332	Aron Baynes	BOS	C	0.17	3.02
333	Grayson Allen	UTA	SG	0.26	3.00
334	Malik Beasley	DEN	SG	0.16	2.99
335	Dante Exum	UTA	PG	0.17	2.97
336	Shelvin Mack	CHA	PG	0.30	2.91
337	Jaron Blossomgame	CLE	SF	0.24	2.89
338	Michael Carter-Williams	HOU	PG	0.33	2.88
339	Nikola Mirotic	MIL	PF	0.19	2.86
340	Vince Carter	ATL	PF	0.17	2.85
341	CJ Miles	MEM	SF	0.39	2.83
342	Tim Frazier	MIL	PG	0.34	2.83
343	Matthew Dellavedova	MIL	PG	0.34	2.81
344	Garrett Temple	LAC	SG	0.29	2.80
345	Alfonzo McKinnie	GSW	SF	0.23	2.80
346	Wayne Ellington	MIA	SG	0.30	2.79
347	Marcin Gortat	LAC	C	0.15	2.77
348	Mike Muscala	LAL	PF	0.29	2.69
349	Langston Galloway	DET	SG	0.16	2.67
350	Rodney Hood	POR	SG	0.25	2.66

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
351	Justin Jackson	DAL	SF	0.35	2.59
352	Mike Scott	PHI	PF	0.35	2.58
353	Henry Ellenson	DET	PF	0.36	2.57
354	Wayne Selden	MEM	SG	0.30	2.53
355	Danuel House Jr.	HOU	SF	0.13	2.53
356	Jonas Jerebko	GSW	PF	0.17	2.51
357	Marquese Chriss	HOU	PF	0.36	2.49
358	Jaylen Adams	ATL	PG	0.22	2.49
359	Brandon Knight	HOU	PG	0.29	2.47
360	Pau Gasol	MIL	C	0.33	2.46
361	Anthony Tolliver	MIN	PF	0.19	2.37
362	Landry Shamet	LAC	SG	0.27	2.35
363	Markieff Morris	OKC	PF	0.21	2.35
364	Khyri Thomas	DET	SG	0.37	2.26
365	Kyle O'Quinn	IND	C	0.20	2.26
366	Timothe Luwawu-Cabarrot	OKC	SF	0.34	2.26
367	Ryan Broekhoff	DAL	SG	0.25	2.24
368	D.J. Wilson	MIL	PF	0.13	2.24
369	Thabo Sefolosha	UTA	SF	0.21	2.22
370	Julian Washburn	MEM	SF	0.20	2.19
371	Wilson Chandler	LAC	PF	0.23	2.13
372	Wesley Matthews	NYK	SF	0.26	2.11
373	Cameron Reynolds	MIN	SG	0.20	2.11
374	Boban Marjanovic	LAC	C	0.23	2.08
375	Tyler Dorsey	ATL	SG	0.29	2.08
376	Austin Rivers	WAS	SG	0.21	2.06
377	Amir Johnson	PHI	C	0.16	2.04
378	Troy Daniels	PHX	SG	0.16	1.91
379	Corey Brewer	PHI	SF	0.25	1.85
380	Deyonta Davis	ATL	C	0.13	1.83
381	Troy Williams	SAC	SF	0.14	1.81

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
382	Zaza Pachulia	DET	C	0.11	1.75
383	Ivica Zubac	LAL	C	0.15	1.75
384	Christian Wood	MIL	PF	0.21	1.73
385	Channing Frye	CLE	C	0.19	1.71
386	Chimezie Metu	SAS	PF	0.26	1.70
387	John Jenkins	WAS	SG	0.33	1.64
388	Thon Maker	MIL	C	0.22	1.63
389	Tyson Chandler	PHX	C	0.20	1.58
390	Milos Teodosic	LAC	PG	0.14	1.58
391	Jonathon Simmons	PHI	SG	0.18	1.55
392	Ben McLemore	SAC	SG	0.20	1.51
393	Doug McDermott	IND	SF	0.12	1.51
394	Glenn Robinson III	DET	SF	0.14	1.50
395	Jordan Bell	GSW	C	0.13	1.45
396	Willy Hernangomez	CHA	C	0.07	1.44
397	Raymond Felton	OKC	PG	0.13	1.43
398	Gary Clark	HOU	PF	0.14	1.40
399	Wesley Johnson	WAS	SF	0.24	1.37
400	Trey Lyles	DEN	PF	0.07	1.35
401	Courtney Lee	NYK	SG	0.25	1.31
402	Guerschon Yabusele	BOS	PF	0.19	1.30
403	Sterling Brown	MIL	SG	0.08	1.29
404	Caleb Swanigan	SAC	PF	0.25	1.28
405	Ryan Anderson	MIA	PF	0.25	1.28
406	Ekpe Udoh	UTA	C	0.16	1.13
407	Daniel Theis	BOS	C	0.07	1.08
408	Iman Shumpert	HOU	SG	0.12	1.07
409	Raul Neto	UTA	PG	0.08	1.05
410	Edmond Sumner	IND	PG	0.11	1.04
411	Deng Adel	CLE	SF	0.15	1.03
412	Aaron Holiday	IND	PG	0.07	0.97

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
413	Brad Wanamaker	BOS	PG	0.11	0.96
414	Kyle Korver	CLE	SG	0.12	0.95
415	Isaiah Canaan	MIL	PG	0.16	0.92
416	MarShon Brooks	MEM	SG	0.06	0.83
417	Jarell Martin	ORL	PF	0.11	0.82
418	Theo Pinson	BKN	SG	0.06	0.73
419	Isaac Bonga	LAL	PG	0.11	0.73
420	Jason Smith	NOP	PF	0.19	0.70
421	Sam Dekker	CLE	PF	0.09	0.70
422	Andrew Bogut	GSW	C	0.05	0.66
423	Daniel Hamilton	ATL	SG	0.05	0.59
424	Jonah Bolden	PHI	PF	0.04	0.59
425	Jared Terrell	MIN	SG	0.08	0.57
426	Chris Boucher	TOR	PF	0.07	0.55
427	Damian Jones	GSW	C	0.03	0.52
428	Deonte Burton	OKC	SG	0.06	0.48
429	Kenneth Faried	BKN	C	0.04	0.47
430	Abdel Nader	OKC	SF	0.04	0.45
431	Semi Ojeleye	BOS	PF	0.05	0.45
432	Jodie Meeks	TOR	SG	0.06	0.42
433	Alex Abrines	OKC	SG	0.03	0.42
434	Juancho Hernangomez	DEN	PF	0.03	0.42
435	Greg Monroe	BOS	C	0.08	0.41
436	Jake Layman	POR	SF	0.02	0.33
437	Sindarius Thornwell	LAC	SG	0.06	0.30
438	Svi Mykhailiuk	DET	SF	0.07	0.29
439	Drew Eubanks	SAS	PF	0.05	0.27
440	Lorenzo Brown	TOR	PG	0.03	0.26
441	Isaiah Hartenstein	HOU	PF	0.03	0.26
442	T.J. McConnell	PHI	PG	0.01	0.25
443	Quincy Pondexter	SAS	SF	0.04	0.22

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
444	Andrew Harrison	MEM	PG	0.05	0.18
445	Alec Burks	SAC	SG	0.02	0.17
446	Tony Snell	MIL	SF	0.01	0.08
447	Anfernee Simons	POR	SG	0.01	0.08
448	Jacob Evans	GSW	SG	0.01	0.08
449	Skal Labissiere	POR	PF	0.02	0.08
450	Patrick Patterson	OKC	PF	0.01	0.07
451	Hamidou Diallo	OKC	SG	0.01	0.07
452	Dante Cunningham	SAS	PF	0.01	0.06

Tabella B.4: Classifica di tutti i giocatori di NBA selezionati dal modello logistico stimato con metodi bayesiani presentati nel Capitolo 3

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
1	Giannis Antetokounmpo	MIL	PF	0.38	22.66
2	James Harden	HOU	PG	0.36	22.52
3	Anthony Davis	NOP	C	0.42	22.16
4	Nikola Jokic	DEN	C	0.40	20.11
5	Russell Westbrook	OKC	PG	0.32	19.36
6	Luka Doncic	DAL	SG	0.40	18.96
7	LeBron James	LAL	SF	0.35	18.95
8	Joel Embiid	PHI	C	0.31	18.36
9	Jrue Holiday	NOP	SG	0.39	17.37
10	Trae Young	ATL	PG	0.40	17.24
11	Karl-Anthony Towns	MIN	C	0.33	17.20
12	Paul George	OKC	SF	0.34	17.17
13	Kawhi Leonard	TOR	SF	0.36	17.01
14	Bradley Beal	WAS	SG	0.37	16.54
15	Jusuf Nurkic	POR	C	0.39	16.32
16	Stephen Curry	GSW	PG	0.36	16.18

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
17	Buddy Hield	SAC	SG	0.47	16.09
18	Kemba Walker	CHA	PG	0.36	15.97
19	Kevin Durant	GSW	SF	0.35	15.80
20	DeMar DeRozan	SAS	SG	0.36	15.71
21	Blake Griffin	DET	PF	0.33	15.68
22	Kevin Love	CLE	PF	0.39	15.18
23	Mike Conley	MEM	PG	0.38	15.13
24	Damian Lillard	POR	PG	0.33	14.96
25	Ben Simmons	PHI	PG	0.34	14.87
26	Devin Booker	PHX	SG	0.30	14.57
27	De'Aaron Fox	SAC	PG	0.36	14.37
28	Nikola Vucevic	ORL	C	0.32	14.35
29	Kyle Lowry	TOR	PG	0.38	14.33
30	Lonzo Ball	LAL	PG	0.46	14.00
31	Julius Randle	NOP	PF	0.31	13.90
32	Hassan Whiteside	MIA	C	0.39	13.61
33	Zach LaVine	CHI	SG	0.32	13.57
34	Pascal Siakam	TOR	PF	0.38	13.51
35	Donovan Mitchell	UTA	SG	0.30	13.37
36	Andre Drummond	DET	C	0.28	13.16
37	DeMarcus Cousins	GSW	C	0.33	12.95
38	Chris Paul	HOU	PG	0.32	12.91
39	Lauri Markkanen	CHI	PF	0.34	12.80
40	Brandon Ingram	LAL	SF	0.35	12.77
41	Klay Thompson	GSW	SG	0.36	12.65
42	Justise Winslow	MIA	SF	0.39	12.53
43	John Wall	WAS	PG	0.27	12.45
44	Danilo Gallinari	LAC	SF	0.35	12.41
45	Dennis Schroder	OKC	PG	0.39	12.39
46	Kyle Kuzma	LAL	PF	0.36	12.39
47	Serge Ibaka	TOR	C	0.39	12.31

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
48	Kyrie Irving	BOS	PG	0.28	12.31
49	Derrick Favors	UTA	C	0.41	12.16
50	Elfrid Payton	NOP	PG	0.37	12.13
51	Deandre Ayton	PHX	C	0.32	12.09
52	Thaddeus Young	IND	PF	0.40	12.00
53	John Collins	ATL	PF	0.29	11.89
54	Ricky Rubio	UTA	PG	0.35	11.89
55	CJ McCollum	POR	SG	0.32	11.86
56	Victor Oladipo	IND	SG	0.30	11.82
57	Tristan Thompson	CLE	C	0.38	11.81
58	Willie Cauley-Stein	SAC	C	0.37	11.81
59	Draymond Green	GSW	PF	0.32	11.78
60	Andrew Wiggins	MIN	SF	0.33	11.71
61	Jeremy Lamb	CHA	SG	0.39	11.70
62	Clint Capela	HOU	C	0.31	11.69
63	Jamal Murray	DEN	PG	0.32	11.68
64	Al Horford	BOS	C	0.35	11.39
65	Taurean Prince	ATL	SF	0.41	11.38
66	Josh Richardson	MIA	SG	0.33	11.28
67	Khris Middleton	MIL	SF	0.30	11.28
68	Rudy Gobert	UTA	C	0.27	11.19
69	Domantas Sabonis	IND	C	0.33	11.14
70	Goran Dragic	MIA	PG	0.36	11.12
71	Jarrett Allen	BKN	C	0.37	10.99
72	Walter Lemon Jr.	CHI	PG	0.32	10.83
73	Jeff Teague	MIN	PG	0.33	10.81
74	Dwyane Wade	MIA	SG	0.34	10.72
75	Jaylen Brown	BOS	SG	0.39	10.67
76	Paul Millsap	DEN	PF	0.34	10.65
77	Rajon Rondo	LAL	PG	0.33	10.62
78	Caris LeVert	BKN	SF	0.34	10.59

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
79	Joe Ingles	UTA	PF	0.34	10.58
80	Lou Williams	LAC	SG	0.28	10.56
81	Steven Adams	OKC	C	0.31	10.53
82	LaMarcus Aldridge	SAS	C	0.25	10.53
83	Tomas Satoransky	WAS	PG	0.42	10.33
84	JaKarr Sampson	CHI	SF	0.29	10.22
85	Marcus Morris	BOS	PF	0.35	10.08
86	JaVale McGee	LAL	C	0.33	9.97
87	Rudy Gay	SAS	PF	0.32	9.91
88	Eric Bledsoe	MIL	PG	0.28	9.89
89	Collin Sexton	CLE	PG	0.32	9.84
90	Jerami Grant	OKC	PF	0.36	9.72
91	Derrick Rose	MIN	PG	0.32	9.72
92	Malcolm Brogdon	MIL	SG	0.32	9.64
93	Marvin Bagley III	SAC	PF	0.30	9.51
94	Shabazz Napier	BKN	PG	0.48	9.50
95	Aaron Gordon	ORL	PF	0.25	9.32
96	D'Angelo Russell	BKN	PG	0.22	9.28
97	Kris Dunn	CHI	PG	0.28	9.27
98	Jayson Tatum	BOS	SF	0.28	9.18
99	Kadeem Allen	NYK	SG	0.38	9.15
100	Jeff Green	WAS	PF	0.38	9.11
101	Larry Nance Jr.	CLE	C	0.31	9.11
102	Bojan Bogdanovic	IND	SF	0.30	9.09
103	Bam Adebayo	MIA	C	0.34	9.08
104	Josh Jackson	PHX	SG	0.31	9.05
105	Alex Len	ATL	C	0.34	9.02
106	Gary Harris	DEN	SG	0.35	8.96
107	Emmanuel Mudiay	NYK	PG	0.29	8.95
108	Thomas Bryant	WAS	C	0.39	8.93
109	Brook Lopez	MIL	C	0.31	8.92

Continua nella prossima pagina

Tabella dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
110	Tony Parker	CHA	PG	0.43	8.85
111	J.J. Barea	DAL	PG	0.33	8.85
112	Dwight Powell	DAL	C	0.37	8.85
113	Darren Collison	IND	PG	0.32	8.82
114	Jae Crowder	UTA	SF	0.34	8.78
115	Jordan Clarkson	CLE	SG	0.31	8.76
116	Montrezl Harrell	LAC	C	0.26	8.70
117	Reggie Jackson	DET	PG	0.29	8.70
118	Patrick Beverley	LAC	PG	0.34	8.66
119	D.J. Augustin	ORL	PG	0.34	8.62
120	Jaren Jackson Jr.	MEM	PF	0.27	8.56
121	Bogdan Bogdanovic	SAC	SG	0.29	8.39
122	Jemerrio Jones	LAL	SF	0.38	8.38
123	Rodions Kurucs	BKN	SF	0.41	8.38
124	Kevin Knox	NYK	PF	0.30	8.17
125	Kyle Anderson	MEM	SF	0.31	8.14
126	Derrick White	SAS	PG	0.31	8.10
127	Marcus Smart	BOS	SG	0.32	8.09
128	Nicolas Batum	CHA	SF	0.33	8.07
129	Mikal Bridges	PHX	SF	0.37	8.04
130	Fred VanVleet	TOR	PG	0.35	8.04
131	PJ Tucker	HOU	PF	0.34	7.98
132	Cody Zeller	CHA	C	0.27	7.96
133	Rondae Hollis-Jefferson	BKN	SF	0.32	7.90
134	Eric Gordon	HOU	SG	0.28	7.87
135	Darius Miller	NOP	SF	0.43	7.80
136	OG Anunoby	TOR	SF	0.47	7.79
137	Myles Turner	IND	C	0.24	7.71
138	Mitchell Robinson	NYK	C	0.33	7.67
139	Dwight Howard	WAS	C	0.23	7.60
140	Joakim Noah	MEM	C	0.33	7.59

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
141	Nemanja Bjelica	SAC	PF	0.31	7.59
142	Johnathan Williams	LAL	C	0.44	7.55
143	Tobias Harris	PHI	PF	0.40	7.53
144	Jimmy Butler	MIN	SF	0.39	7.52
145	Allonzo Trier	NYK	SG	0.33	7.48
146	Mason Plumlee	DEN	C	0.29	7.48
147	JJ Redick	PHI	SG	0.25	7.42
148	Gordon Hayward	BOS	PF	0.29	7.38
149	Alex Caruso	LAL	PG	0.31	7.37
150	Monte Morris	DEN	PG	0.38	7.37
151	Derrick Jones Jr.	MIA	SF	0.40	7.36
152	Miles Bridges	CHA	SF	0.40	7.36
153	Dion Waiters	MIA	SG	0.31	7.35
154	Tyreke Evans	IND	SG	0.30	7.32
155	Taj Gibson	MIN	PF	0.28	7.32
156	Spencer Dinwiddie	BKN	PG	0.23	7.26
157	E'Twaun Moore	NOP	SG	0.33	7.26
158	Tyus Jones	MIN	PG	0.37	7.25
159	Kentavious Caldwell-Pope	LAL	SG	0.35	7.25
160	Noah Vonleh	NYK	PF	0.27	7.23
161	Jahlil Okafor	NOP	C	0.36	7.22
162	Al-Farouq Aminu	POR	PF	0.29	7.21
163	Enes Kanter	POR	C	0.44	7.13
164	Shai Gilgeous-Alexander	LAC	PG	0.27	7.10
165	Josh Okogie	MIN	SG	0.33	7.08
166	Kent Bazemore	ATL	SG	0.25	7.08
167	De'Anthony Melton	PHX	PG	0.35	7.02
168	Ryan Arcidiacono	CHI	PG	0.38	6.98
169	Will Barton	DEN	SF	0.27	6.94
170	Maurice Harkless	POR	SF	0.31	6.92
171	Terrence Ross	ORL	SG	0.27	6.90

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
172	Deyonta Davis	ATL	C	0.49	6.86
173	Richaun Holmes	PHX	C	0.32	6.85
174	Frank Kaminsky	CHA	C	0.36	6.83
175	Willy Hernangomez	CHA	C	0.35	6.82
176	Andre Iguodala	GSW	SF	0.34	6.79
177	Bruno Caboclo	MEM	SF	0.31	6.77
178	Allen Crabbe	BKN	SG	0.32	6.77
179	Evan Fournier	ORL	SG	0.22	6.77
180	Dorian Finney-Smith	DAL	SF	0.32	6.75
181	Bryn Forbes	SAS	SG	0.29	6.66
182	Jamal Crawford	PHX	SG	0.35	6.65
183	Josh Hart	LAL	SG	0.33	6.61
184	Kelly Olynyk	MIA	PF	0.28	6.59
185	Markelle Fultz	PHI	SG	0.26	6.59
186	T.J. Warren	PHX	SF	0.21	6.53
187	Evan Turner	POR	PG	0.32	6.52
188	Treveon Graham	BKN	SG	0.43	6.49
189	Kevin Huerter	ATL	SG	0.27	6.46
190	Kenrich Williams	NOP	SF	0.33	6.44
191	Omri Casspi	MEM	SF	0.45	6.41
192	Dragan Bender	PHX	PF	0.37	6.41
193	Jose Calderon	DET	PG	0.62	6.40
194	Wes Iwundu	ORL	SF	0.45	6.34
195	Marc Gasol	TOR	C	0.37	6.33
196	James Johnson	MIA	PF	0.31	6.33
197	Harry Giles III	SAC	PF	0.32	6.23
198	Jalen Brunson	DAL	PG	0.30	6.22
199	Patty Mills	SAS	PG	0.31	6.19
200	Malik Monk	CHA	SG	0.34	6.18
201	Bonzie Colson	MIL	SF	0.42	6.11
202	DeAndre' Bembry	ATL	SG	0.25	6.10

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
203	Royce O'Neale	UTA	SF	0.35	6.05
204	Danny Green	TOR	SG	0.29	6.03
205	Edmond Sumner	IND	PG	0.65	6.02
206	Chandler Parsons	MEM	SF	0.30	6.00
207	Mohamed Bamba	ORL	C	0.32	5.99
208	Wendell Carter Jr.	CHI	C	0.20	5.97
209	Torrey Craig	DEN	SF	0.35	5.95
210	Mario Hezonja	NYK	SF	0.27	5.90
211	Joe Harris	BKN	SG	0.23	5.89
212	Ersan Ilyasova	MIL	PF	0.31	5.89
213	Ante Zizic	CLE	C	0.29	5.86
214	Marco Belinelli	SAS	SG	0.30	5.82
215	Devin Harris	DAL	PG	0.36	5.81
216	Terry Rozier	BOS	PG	0.27	5.76
217	Maxi Kleber	DAL	PF	0.30	5.76
218	Dewayne Dedmon	ATL	C	0.19	5.75
219	Salah Mejri	DAL	C	0.40	5.69
220	Dwayne Bacon	CHA	SG	0.35	5.55
221	Robin Lopez	CHI	C	0.26	5.52
222	Luke Kennard	DET	SG	0.29	5.51
223	DeAndre Jordan	NYK	C	0.31	5.45
224	Luke Kornet	NYK	PF	0.32	5.43
225	Frank Ntilikina	NYK	PG	0.28	5.41
226	Lorenzo Brown	TOR	PG	0.62	5.36
227	Nerlens Noel	OKC	C	0.33	5.33
228	Antonio Blakeney	CHI	SG	0.36	5.26
229	Dario Saric	PHI	PF	0.43	5.23
230	Elie Okobo	PHX	PG	0.31	5.15
231	Trey Lyles	DEN	PF	0.26	5.13
232	Cedi Osman	CLE	SF	0.18	5.12
233	Lance Stephenson	LAL	SG	0.28	5.10

Continua nella prossima pagina

Tabella dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
234	Dirk Nowitzki	DAL	PF	0.29	5.09
235	Damyean Dotson	NYK	SG	0.23	5.09
236	Tim Hardaway Jr.	DAL	SG	0.31	5.09
237	Robert Covington	PHI	SF	0.32	5.09
238	Jevon Carter	MEM	PG	0.36	5.08
239	Cheick Diallo	NOP	PF	0.30	5.06
240	Michael Kidd-Gilchrist	CHA	PF	0.27	5.04
241	Trevor Ariza	PHX	SF	0.34	5.00
242	Ian Clark	NOP	SG	0.34	4.97
243	Isaiah Thomas	DEN	SG	0.27	4.97
244	Marvin Williams	CHA	PF	0.21	4.95
245	Kevon Looney	GSW	C	0.26	4.94
246	Gorgui Dieng	MIN	C	0.28	4.90
247	Omari Spellman	ATL	PF	0.28	4.89
248	Frank Jackson	NOP	PG	0.30	4.88
249	Carmelo Anthony	HOU	PF	0.18	4.88
250	Sterling Brown	MIL	SG	0.31	4.87
251	Norman Powell	TOR	SG	0.30	4.85
252	Kyle O'Quinn	IND	C	0.42	4.83
253	Bobby Portis	CHI	PF	0.30	4.79
254	Nene	HOU	C	0.39	4.75
255	Dennis Smith Jr.	NYK	PG	0.29	4.71
256	Pat Connaughton	MIL	SG	0.26	4.67
257	Khem Birch	ORL	C	0.33	4.67
258	Juancho Hernangomez	DEN	PF	0.34	4.67
259	Danuel House Jr.	HOU	SF	0.25	4.67
260	Abdel Nader	OKC	SF	0.45	4.61
261	Cristiano Felicio	CHI	C	0.38	4.60
262	Dante Exum	UTA	PG	0.26	4.60
263	D.J. Wilson	MIL	PF	0.28	4.59
264	Jonathan Isaac	ORL	PF	0.19	4.59

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
265	Alex Poythress	ATL	PF	0.30	4.58
266	Jonas Valanciunas	TOR	C	0.26	4.58
267	Kelly Oubre Jr.	WAS	SF	0.31	4.57
268	Ed Davis	BKN	C	0.21	4.53
269	Rodney McGruder	MIA	SG	0.23	4.52
270	Damian Jones	GSW	C	0.29	4.51
271	Robert Williams III	BOS	C	0.45	4.49
272	Mo Bamba	ORL	C	0.23	4.48
273	Ish Smith	DET	PG	0.21	4.48
274	Shaquille Harrison	CHI	SG	0.24	4.45
275	Devonte' Graham	CHA	PG	0.32	4.39
276	Delon Wright	TOR	PG	0.39	4.37
277	John Henson	MIL	C	0.25	4.30
278	Meyers Leonard	POR	C	0.29	4.30
279	Tyrone Wallace	LAC	PG	0.39	4.29
280	Jordan Bell	GSW	C	0.38	4.27
281	Davis Bertans	SAS	PF	0.25	4.27
282	Jared Dudley	BKN	PF	0.29	4.26
283	Jabari Parker	WAS	PF	0.27	4.25
284	Ivan Rabb	MEM	PF	0.25	4.25
285	Otto Porter Jr.	CHI	SF	0.30	4.22
286	Rawle Alkins	CHI	SG	0.36	4.19
287	Cory Joseph	IND	PG	0.21	4.16
288	Jordan McRae	WAS	SG	0.32	4.15
289	Dillon Brooks	MEM	SF	0.23	4.14
290	Bismack Biyombo	CHA	C	0.26	4.13
291	Jerian Grant	ORL	PG	0.32	4.11
292	Aaron Holiday	IND	PG	0.31	4.11
293	Damion Lee	GSW	SG	0.42	4.10
294	Harrison Barnes	SAC	PF	0.29	4.09
295	Aron Baynes	BOS	C	0.23	4.08

Continua nella prossima pagina

Tabella dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
296	Amir Johnson	PHI	C	0.31	4.06
297	DeMarre Carroll	BKN	PF	0.17	4.04
298	Alec Burks	SAC	SG	0.58	4.04
299	Johnathan Motley	LAC	PF	0.37	4.03
300	Moritz Wagner	LAL	C	0.33	4.02
301	Jakob Poeltl	SAS	C	0.22	4.00
302	Yogi Ferrell	SAC	PG	0.31	4.00
303	Cameron Payne	CLE	PG	0.49	4.00
304	Zach Collins	POR	C	0.21	3.96
305	Donte DiVincenzo	MIL	SG	0.30	3.95
306	Miles Plumlee	ATL	C	0.34	3.94
307	Lonnie Walker IV	SAS	SG	0.59	3.85
308	Avery Bradley	MEM	SG	0.33	3.84
309	Trey Burke	DAL	PG	0.37	3.80
310	Kosta Koufos	SAC	C	0.26	3.80
311	Jonas Jerebko	GSW	PF	0.25	3.79
312	Lance Thomas	NYK	PF	0.29	3.79
313	Keita Bates-Diop	MIN	SF	0.28	3.79
314	JaMychal Green	LAC	PF	0.32	3.79
315	Troy Brown Jr.	WAS	SF	0.28	3.77
316	Michael Beasley	LAL	PF	0.24	3.75
317	Luol Deng	MIN	SF	0.24	3.74
318	Chandler Hutchison	CHI	SF	0.25	3.71
319	Ray Spalding	PHX	PF	0.25	3.70
320	Quinn Cook	GSW	PG	0.28	3.69
321	TJ Leaf	IND	PF	0.42	3.69
322	Gerald Green	HOU	SG	0.23	3.67
323	Duncan Robinson	MIA	PF	0.42	3.66
324	Malik Beasley	DEN	SG	0.20	3.65
325	Jerome Robinson	LAC	SG	0.41	3.65
326	Yuta Watanabe	MEM	SF	0.39	3.61

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
327	Shelvin Mack	CHA	PG	0.37	3.61
328	Furkan Korkmaz	PHI	SG	0.27	3.60
329	David Nwaba	CLE	SG	0.21	3.59
330	Tim Frazier	MIL	PG	0.44	3.59
331	Justin Anderson	ATL	SF	0.33	3.58
332	Isaiah Briscoe	ORL	PG	0.28	3.55
333	Markieff Morris	OKC	PF	0.32	3.53
334	Tyler Johnson	PHX	PG	0.29	3.51
335	Jerryd Bayless	MIN	PG	0.19	3.50
336	Jalen Jones	CLE	SF	0.27	3.50
337	Solomon Hill	NOP	SF	0.24	3.49
338	Frank Mason	SAC	PG	0.26	3.48
339	Jaron Blossomgame	CLE	SF	0.29	3.47
340	Bruce Brown	DET	SG	0.23	3.46
341	Brandon Sampson	CHI	SG	0.30	3.44
342	Terrance Ferguson	OKC	SG	0.20	3.42
343	Justin Holiday	CHI	SG	0.28	3.41
344	Chasson Randle	WAS	PG	0.25	3.41
345	Nik Stauskas	CLE	SG	0.54	3.39
346	Grayson Allen	UTA	SG	0.30	3.35
347	Quincy Acy	PHX	PF	0.29	3.35
348	Jeremy Lin	TOR	PG	0.32	3.32
349	Shake Milton	PHI	SG	0.29	3.30
350	JR Smith	CLE	SG	0.18	3.27
351	Tyler Zeller	ATL	C	0.34	3.27
352	Boban Marjanovic	LAC	C	0.36	3.27
353	Marcin Gortat	LAC	C	0.18	3.25
354	Seth Curry	POR	SG	0.23	3.23
355	Ian Mahinmi	WAS	C	0.21	3.22
356	George Hill	CLE	PG	0.33	3.21
357	Stanley Johnson	NOP	SF	0.35	3.20

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
358	Tony Snell	MIL	SF	0.26	3.15
359	Alfonzo McKinnie	GSW	SF	0.26	3.11
360	Raymond Felton	OKC	PG	0.28	3.11
361	Jon Leuer	DET	PF	0.31	3.07
362	Reggie Bullock	LAL	SG	0.29	3.07
363	Georges Niang	UTA	PF	0.34	3.04
364	T.J. McConnell	PHI	PG	0.18	2.98
365	Alex Abrines	OKC	SG	0.23	2.95
366	James Ennis III	PHI	SF	0.32	2.91
367	Quincy Pondexter	SAS	SF	0.49	2.84
368	MarShon Brooks	MEM	SG	0.21	2.84
369	Michael Carter-Williams	HOU	PG	0.33	2.83
370	Brad Wanamaker	BOS	PG	0.33	2.82
371	Daniel Hamilton	ATL	SG	0.24	2.77
372	Mike Muscala	LAL	PF	0.30	2.76
373	Matthew Dellavedova	MIL	PG	0.33	2.71
374	Zaza Pachulia	DET	C	0.17	2.70
375	Troy Williams	SAC	SF	0.21	2.69
376	Wayne Ellington	MIA	SG	0.29	2.69
377	Daniel Theis	BOS	C	0.18	2.68
378	Vince Carter	ATL	PF	0.16	2.66
379	CJ Miles	MEM	SF	0.37	2.65
380	Andrew Bogut	GSW	C	0.19	2.65
381	Garrett Temple	LAC	SG	0.27	2.63
382	Nikola Mirotic	MIL	PF	0.18	2.63
383	Jake Layman	POR	SF	0.17	2.62
384	Anthony Tolliver	MIN	PF	0.20	2.60
385	Patrick McCaw	CLE	SG	0.64	2.57
386	Hamidou Diallo	OKC	SG	0.22	2.57
387	Langston Galloway	DET	SG	0.16	2.52
388	Rodney Hood	POR	SG	0.24	2.52

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
389	Pau Gasol	MIL	C	0.33	2.51
390	Isaiah Hartenstein	HOU	PF	0.27	2.51
391	Wayne Selden	MEM	SG	0.29	2.50
392	Jaylen Adams	ATL	PG	0.22	2.49
393	Gary Clark	HOU	PF	0.26	2.48
394	Justin Jackson	DAL	SF	0.33	2.48
395	Theo Pinson	BKN	SG	0.22	2.47
396	Ivica Zubac	LAL	C	0.22	2.47
397	Sindarius Thornwell	LAC	SG	0.46	2.46
398	Shaun Livingston	GSW	PG	0.20	2.46
399	Mike Scott	PHI	PF	0.33	2.45
400	Jarell Martin	ORL	PF	0.33	2.44
401	Brandon Knight	HOU	PG	0.29	2.43
402	Kenneth Faried	BKN	C	0.22	2.43
403	Raul Neto	UTA	PG	0.18	2.40
404	Channing Frye	CLE	C	0.26	2.40
405	Khyri Thomas	DET	SG	0.39	2.40
406	Henry Ellenson	DET	PF	0.34	2.37
407	Marquese Chriss	HOU	PF	0.34	2.35
408	Wilson Chandler	LAC	PF	0.26	2.35
409	Julian Washburn	MEM	SF	0.21	2.34
410	Ryan Broekhoff	DAL	SG	0.25	2.30
411	Landry Shamet	LAC	SG	0.26	2.29
412	Cameron Reynolds	MIN	SG	0.21	2.25
413	Timothe Luwawu-Cabarrot	OKC	SF	0.33	2.23
414	Tyson Chandler	PHX	C	0.27	2.15
415	Ben McLemore	SAC	SG	0.28	2.14
416	Jonah Bolden	PHI	PF	0.14	2.10
417	Ekpe Udoh	UTA	C	0.29	2.09
418	Iman Shumpert	HOU	SG	0.23	2.06
419	Semi Ojeleye	BOS	PF	0.25	2.05

Continua nella prossima pagina

Tabelle dettagliate

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
420	Troy Daniels	PHX	SG	0.17	2.01
421	Wesley Matthews	NYK	SF	0.25	2.01
422	Tyler Dorsey	ATL	SG	0.28	2.01
423	Dante Cunningham	SAS	PF	0.19	1.96
424	Chimezie Metu	SAS	PF	0.29	1.95
425	Thabo Sefolosha	UTA	SF	0.19	1.95
426	Austin Rivers	WAS	SG	0.20	1.94
427	Glenn Robinson III	DET	SF	0.17	1.94
428	Ryan Anderson	MIA	PF	0.38	1.94
429	Doug McDermott	IND	SF	0.15	1.93
430	Deonte Burton	OKC	SG	0.24	1.79
431	Christian Wood	MIL	PF	0.21	1.76
432	Corey Brewer	PHI	SF	0.23	1.75
433	Milos Teodosic	LAC	PG	0.15	1.74
434	Jonathon Simmons	PHI	SG	0.20	1.68
435	Patrick Patterson	OKC	PF	0.16	1.66
436	Thon Maker	MIL	C	0.22	1.65
437	Caleb Swanigan	SAC	PF	0.31	1.59
438	Deng Adel	CLE	SF	0.23	1.57
439	John Jenkins	WAS	SG	0.32	1.57
440	Guerschon Yabusele	BOS	PF	0.22	1.53
441	Wesley Johnson	WAS	SF	0.26	1.45
442	Kyle Korver	CLE	SG	0.17	1.43
443	Isaac Bonga	LAL	PG	0.21	1.42
444	Chris Boucher	TOR	PF	0.17	1.37
445	Courtney Lee	NYK	SG	0.25	1.32
446	Anfernee Simons	POR	SG	0.20	1.26
447	Jodie Meeks	TOR	SG	0.18	1.23
448	Jared Terrell	MIN	SG	0.16	1.17
449	Jacob Evans	GSW	SG	0.19	1.09
450	Isaiah Canaan	MIL	PG	0.19	1.06

Continua nella prossima pagina

	Giocatore	Squadra	Ruolo	Ranking	Ranking corretto
451	Svi Mykhailiuk	DET	SF	0.20	0.84
452	Jason Smith	NOP	PF	0.23	0.83

Tabella B.5: Migliori 10 giocatori di ruolo alla grande classificati dal modello gerarchico stimato con metodi bayesiani

	Giocatore	Squadra	Ranking	Ranking corretto
	Giannis Antetokounmpo	MIL	0.38	22.83
	Blake Griffin	DET	0.32	14.96
	Julius Randle	NOP	0.33	14.64
	John Collins	ATL	0.33	13.50
	Kevin Love	CLE	0.35	13.48
	Pascal Siakam	TOR	0.36	12.95
	Lauri Markkanen	CHI	0.34	12.64
	Draymond Green	GSW	0.33	11.88
	Kyle Kuzma	LAL	0.33	11.40
	Aaron Gordon	ORL	0.29	10.58

Tabella B.6: Migliori 10 giocatori di ruolo centro classificati dal modello gerarchico stimato con metodi bayesiani

Giocatore	Squadra	Ranking	Ranking corretto
Joel Embiid	PHI	0.33	19.83
Nikola Jokic	DEN	0.39	19.65
Anthony Davis	NOP	0.37	19.56
Karl-Anthony Towns	MIN	0.34	17.58
Jusuf Nurkic	POR	0.35	14.95
DeMarcus Cousins	GSW	0.37	14.52
Andre Drummond	DET	0.27	13.05
Rudy Gobert	UTA	0.31	12.73
LaMarcus Aldridge	SAS	0.30	12.66
Nikola Vucevic	ORL	0.28	12.39

Tabella B.7: Migliori 10 giocatori di ruolo playmaker classificati dal modello gerarchico stimato con metodi bayesiani

Giocatore	Squadra	Ranking	Ranking corretto
James Harden	HOU	0.36	22.46
Russell Westbrook	OKC	0.33	19.90
Kemba Walker	CHA	0.38	16.80
Trae Young	ATL	0.39	16.79
Stephen Curry	GSW	0.37	16.51
Mike Conley	MEM	0.40	16.16
De'Aaron Fox	SAC	0.41	16.11
Damian Lillard	POR	0.34	15.31
John Wall	WAS	0.33	15.21
Ben Simmons	PHI	0.33	14.57

Tabella B.8: Migliori 10 giocatori di ruolo guardia tiratrice classificati dal modello gerarchico stimato con metodi bayesiani

Giocatore	Squadra	Ranking	Ranking corretto
Luka Doncic	DAL	0.40	18.59
Devin Booker	PHX	0.37	17.79
Bradley Beal	WAS	0.38	17.01
DeMar DeRozan	SAS	0.36	15.73
Jrue Holiday	NOP	0.35	15.60
Zach LaVine	CHI	0.35	14.66
Donovan Mitchell	UTA	0.30	13.09
Victor Oladipo	IND	0.32	12.74
Klay Thompson	GSW	0.36	12.44
Jeremy Lamb	CHA	0.39	11.66

Tabella B.9: Migliori 10 giocatori di ruolo ala piccola classificati dal modello gerarchico stimato con metodi bayesiani

Giocatore	Squadra	Ranking	Ranking corretto
LeBron James	LAL	0.38	20.49
Paul George	OKC	0.37	18.70
Kawhi Leonard	TOR	0.38	17.74
Kevin Durant	GSW	0.36	16.32
Brandon Ingram	LAL	0.34	12.32
Andrew Wiggins	MIN	0.34	11.81
Danilo Gallinari	LAC	0.33	11.72
Caris LeVert	BKN	0.37	11.45
Khris Middleton	MIL	0.31	11.43
JaKarr Sampson	CHI	0.32	11.23

Tabelle dettagliate

Ringraziamenti

In queste poche righe voglio ringraziare tutte le persone che sono state presenti per me fino a questo momento.

Tutta la mia famiglia, che mi ha permesso di affrontare al meglio questo percorso di studi, senza mai negarmi nessuna possibilità.

La mia fidanzata Eleonora, che soprattutto nei momenti più difficili mi è stata accanto.

I miei amici, che con la loro spensieratezza mi hanno sempre supportato.

Il Professore Nicola Sartori, e tutti gli altri insegnanti, che sono riusciti a farmi appassionare a questa fantastica scienza che è la statistica.