Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa

RELAZIONE FINALE

Modelli predittivi nel tennis: evoluzione e applicazioni

Relatore Prof. Francesco Lisi
Dipartimento di Scienze Statistiche

Laureando: Francesco Pozza
Matricola N. 2042565

Anno Accademico 2025/2026

Indice

1 Introduzione

La capacità di prevedere correttamente l'esito di un evento sportivo è da sempre una sfida che ha interessato gli appassionati di ogni disciplina. Il processo fondamentale che ha permesso di superare la mera speculazione, trasformando i tentativi di stima da intuizioni di fanatici che provavano alla cieca a indovinare i risultati di un determinato avvenimento ad una vera e propria analisi scrupolosa, è stata l'introduzione di modelli matematici previsionali.

Il tennis non fa eccezione: il progresso nell'affrontare il grattacapo della previsione tennistica è iniziato solo pochi decenni fa con la pubblicazione dei primi modelli matematici per pronosticare l'esito dei match. Supportato dal fatto che il tennis abbia solo due risultati possibili, la vittoria e la sconfitta, i candidati più ovvi per la previsione dei risultati dei match furono modelli di regressione logistica binaria. In aggiunta, l'ecosistema del tennis è unico e affascinante per la modellazione statistica: oltre all'assenza dei pareggi, è uno sport individuale (nel singolare) caratterizzato da un sistema di punteggio gerarchico e non lineare (punti, game, set, match) e influenzato da variabili esterne, come la superficie del terreno di gioco o il formato del torneo.

Da quando l'analisi quantitativa si è affermata nel professionismo, sono stati sviluppati molti altri modelli e l'interesse per gli approcci statistici alla previsione delle vittorie nel tennis ha continuato a proliferare, rendendo quello che un tempo era un campo dominato dall'intuito un terreno fertile per un meticoloso approccio *data-driven*. Nonostante la letteratura offrisse un'ampia proposta riguardante i modelli predittivi, è mancata a lungo una valutazione sistematica e comparativa dell'effettiva efficacia degli stessi. Proprio per colmare questo vuoto, il lavoro proposto da Kovalchik [1], che sarà affrontato nel dettaglio nei prossimi capitoli, ha la finalità di testare le performance di una serie di modelli pubblicati per la previsione degli esiti dei match di singolare nel tennis professionistico, sperimentandoli su un dataset di risultati di match di singolare relativi alla stagione 2014 dell'ATP Tour.

Attraverso il suo studio, l'autrice mira a identificare i principali determinanti della capacità di vittoria nel tennis professionistico e a evidenziare le modalità per migliorare le performance dei modelli attuali, tramite un confronto rigoroso contro un benchmark solido, che in letteratura è rappresentato dal Bookmaker Consensus Model (BCM) [2], ovvero l'aggregazione delle quote espresse dal mercato delle scommesse.

Una volta replicati gli spunti proposti dalla letteratura e conseguentemente avendo la certezza di un *benchmark* solido, il fine ultimo di questo elaborato è quello di testare i modelli in questione sui match del circuito ATP moderno per verificarne la validità nel tempo. L'obiettivo a cui si punta è quello di mettersi nei panni di un appassionato o di uno scommettitore che voglia confrontarsi con il mercato delle scommesse, senza però avere le informazioni in possesso dei bookmakers, e valutare se esista concretamente la

tanto ricercata possibilità di battere il banco unicamente attraverso l'utilizzo di modelli predittivi.

2 Tennis Analytics

In questo capitolo, tramite i riferimenti proposti dalla letteratura, verrà illustrato il progressivo sviluppo e l'applicazione dell'analisi statistica a fini predittivi nell'ambiente tennistico. L'obiettivo è quello di definire lo stato dell'arte, in modo da delineare l'impalcatura metodologica dei modelli che verranno trattati nei capitoli successivi e inquadrarli rispetto all'evoluzione proposta dalla letteratura.

2.1 L'evoluzione della statistica nel tennis

Nell'ultimo decennio, l'analisi statistica relativa al tennis ha subito una forte crescita, dovuta alla possibilità di ottenere una mole di dati estremamente dettagliati, disponibili grazie all'implementazione di nuove tecnologie che stanno spingendo la ricerca verso lo studio sempre più accurato di pattern tattici. Tuttavia, la letteratura scientifica dello scorso decennio si è prettamente dedicata alla previsione dell'esito dell'incontro, basandosi su dati che escludevano le dinamiche dello scambio e che comprendevano unicamente statistiche riassuntive del match.

Dalla letteratura emerge quindi una dicotomia: da un lato i più recenti e notevolmente dettagliati modelli spaziali e dall'altro lato modelli macroscopici, attendibili ma meno minuziosi tatticamente. Come premesso, la letteratura riguardante l'analisi statistica del tennis affonda le proprie radici nei modelli strutturati sui dati riassuntivi dell'incontro, sia per una questione pratica legata alla disponibilità del dato, sia perché risultano essere i più immediati da impostare ed applicare. In questa categoria rientrano la maggioranza dei modelli discussi in letteratura a partire dalle prime pubblicazioni sul tema fino all'introduzione dell'Hawk-Eye, il sistema di tracciamento ottico per la ricostruzione della traiettoria della pallina, che ha spostato l'interesse della ricerca dall'analisi descrittiva all'analisi dinamica del movimento e delle tattiche tramite machine learning. I principali modelli predittivi per stimare le probabilità di vittoria ricadono nelle seguenti tipologie [1]:

- Modelli basati sulla regressione (*Regression-based*);
- Modelli basati sui punti (*Point-based*);
- Modelli basati sul confronto a coppie (*Paired comparison*).

2.2 Modelli basati sulla regressione

I modelli appartenenti a questa categoria stimano direttamente il vincitore del match utilizzando specifiche variabili esplicative. L'approccio più diffuso, nonché uno tra i primi

discussi in letteratura, è quello che usa una regressione probit e il ranking dei giocatori come unico predittore. In un modello probit la probabilità di vittoria è definita come:

$$\pi_{ij} = \Phi(x'_{ij}\beta) \tag{1}$$

dove Φ indica la funzione di ripartizione di una variabile normale standard e x_{ij} rappresenta un vettore di predittori che potrebbero includere le caratteristiche del giocatore, dell'avversario o dell'incontro [1].

Il modello pioniere è il *Basic Probit* di Boulier e Stekler (1999) [3], che si distingue per la sua estrema semplicità: l'unico predittore utilizzato per stimare le probabilità di vittoria è la differenza fra i numeri di testa di serie dei due avversari; viene introdotto anche un indicatore specifico nel caso di sfidanti non classificati tra le teste di serie. Attraverso questo modello gli autori propongono uno dei primi tentativi di formalizzare la previsione tennistica attraverso un modello statistico strutturato, nonostante l'utilizzo della sola differenza tra i numeri di testa di serie come covariata ne limiti le potenzialità.

Il passo successivo, per superare i limiti dell'unica variabile, è il modello logistico proposto da Klaassen e Magnus (2003) [4]. Definendo R_i come il ranking del giocatore i, considerato favorito per la vittoria del match, e R_j come il ranking dell'avversario j, il modello logit da loro proposto è il seguente:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \theta \log\left(\frac{R_i}{R_j}\right) \tag{2}$$

In questa formulazione, l'unica variabile esplicativa del modello è rappresentata dal logaritmo del rapporto tra i ranking dei due giocatori, capace di catturare in modo più continuo la differenza di abilità.

Con l'aumentare della capacità computazionale, sono stati vagliati modelli probit più complessi. Rilevante è il *Probit Plus* di Del Corral e Prieto-Rodriguez (2010) [5], i quali hanno dimostrato come si possa migliorare significativamente l'accuratezza predittiva rispetto ai modelli base attraverso l'incorporazione di ulteriori covariate biometriche e storiche: differenza di età e di altezza, mano dominante, turno di gioco e livello del torneo. Gli autori non si sono limitati unicamente a testare il modello ma ne hanno confrontato tre distinte configurazioni:

- **Modello completo:** include la differenza nel ranking ATP, una variabile dummy per l'appartenenza alla *top ten* in carriera, le caratteristiche biometriche (altezza, età, mano dominante) e variabili strutturali del torneo (turno di gioco, livello del torneo).

- **Modello senza ranking:** esclude le variabili basate sulla classifica.

- **Modello senza caratteristiche fisiche:** esclude le variabili biometriche.

Tale approccio ha permesso di isolare il peso relativo di ogni classe di predittori nella stima della probabilità di vittoria, superando lo scoglio del puro merito sportivo identificato dal ranking.

2.3 Modelli basati sui punti

A differenza dei regressori che operano a livello di match, gli approcci *point-based* scompongono l'incontro nella sua unità fondamentale. Questi modelli stimano inizialmente la probabilità che un giocatore vinca un punto al servizio o in risposta contro uno specifico avversario. I primi autori a spostare la ricerca dalla modellazione del match alla modellazione del singolo punto furono Klaassen e Magnus [6]; dal loro lavoro si evidenzia che l'assunzione matematica fondamentale, e al contempo il principale limite teorico, di questi modelli è che i punti, pur non essendo di fatto indipendenti e identicamente distribuiti perché influenzati dall'inerzia psicologica del match (il *momentum*), possano essere statisticamente approssimati come tali. Tuttavia, offrono il vantaggio inestimabile di poter entrare nel dettaglio della singola partita, prevedendo ad esempio la probabilità di vincere un set o di arrivare al tie-break, superando la dicotomia del solo risultato finale.

Il punto di svolta metodologico per questa categoria di modelli è rappresentato da Newton e Keller [7], i quali hanno fornito una delle prime trattazioni sistematiche del modello *point-based*. Gli autori hanno modellizzato il tennis come una Catena di Markov gerarchica, dimostrando come, partendo dalla probabilità di vincere un singolo punto al servizio, sia possibile ricostruire analiticamente l'intero sistema di punteggio del tennis (game, set e match), trattando il gioco come un processo stocastico a più stadi.

Successivamente al lavoro di Newton e Keller, il dibattito accademico si è concentrato su come stimare nel modo più accurato possibile gli input di base del modello, ovvero le probabilità di vincere un punto al servizio e in risposta.

Barnett e Clarke [8] hanno proposto di calcolare una media specifica del torneo, aggiustata in base al vantaggio del giocatore al servizio e dell'avversario in risposta. Tale vantaggio veniva stimato quantificando di quanto le abilità di un singolo tennista fossero superiori (o inferiori) rispetto a quelle del giocatore medio del circuito.

In seguito, Spanias e Knottenbelt [9] hanno perfezionato questo approccio introducendo un modello "a basso livello" (*low-level point model*). Invece di valutare il punto nella sua interezza, gli autori hanno modellato i singoli stati che conducono alla vittoria del punto stesso (es. *ace*, punto vinto con la prima o con la seconda di servizio). Le probabilità di questi stati, calcolate sulla base dei match recenti e aggiustate per la forza dell'avversario seguendo la logica di Barnett e Clarke, vengono poi combinate per ottenere una stima complessiva. Testando il modello sulla stagione ATP 2011, gli autori hanno dimostrato che l'utilizzo dei dati relativi ai 12 mesi precedenti l'incontro minimizza l'errore predittivo rispetto a finestre temporali più brevi o più lunghe (6 o 18 mesi).

Un'ulteriore e ingegnosa estensione è stata introdotta da Knottenbelt, Spanias e Madurska [10]. Per risolvere il problema delle stime distorte, causato dal fatto che tennisti

diversi affrontano avversari di calibro diverso, gli autori hanno sviluppato un modello "ad avversario comune" (*common-opponent model*). In questa formulazione, le probabilità di servizio e risposta vengono stimate calcolando le statistiche esclusivamente su un sottoinsieme di match giocati contro avversari che entrambi i tennisti oggetto di previsione hanno affrontato in passato. Isolando l'effetto della qualità dell'avversario, questo approccio elimina i bias legati al calendario, fornendo una stima nettamente più equa e bilanciata delle forze in campo.

2.4 Modelli basati sul confronto a coppie

I modelli di confronto a coppie nascono dall'idea di valutare direttamente l'abilità di due contendenti basandosi sull'esito dei loro scontri passati o sulle loro performance contro avversari comuni. L'approccio accademico più celebre di questa categoria è quello proposto da McHale e Morton [11], i quali hanno adattato con successo il modello matematico di Bradley-Terry al tennis. L'assunto di base è che ogni giocatore possegga un'abilità latente di vittoria, definita come α_i. Nella formulazione classica di Bradley-Terry, la probabilità π_{ij} che il giocatore i sconfigga il giocatore j è data dal rapporto della propria abilità sull'abilità totale in campo:

$$\pi_{ij} = \frac{\alpha_i}{\alpha_i + \alpha_j}$$

Un'evoluzione particolarmente nota e performante all'interno dei modelli di confronto a coppie è il sistema Elo. Originariamente concepito per classificare i giocatori di scacchi, questo algoritmo dinamico è stato riadattato al mondo del tennis dal portale statistico Fischer-Baum, Flowers e Bialik [12]. La peculiarità di questo sistema risiede nell'utilizzo di una formula ricorsiva per aggiornare l'abilità latente del giocatore i-esimo immediatamente dopo ogni incontro:

$$E_i(t+1) = E_i(t) + K(t)(W_i(t) - \hat{\pi}_{ij}(t)) \tag{3}$$

In questa equazione di aggiornamento, t indica l'indice temporale del match, $W_i(t)$ è la variabile indicatrice dell'esito (pari a 1 in caso di vittoria, 0 in caso di sconfitta) ed $E_i(t)$ rappresenta il rating Elo del giocatore all'inizio della partita. Il termine $\hat{\pi}_{ij}(t)$ è la probabilità di vittoria attesa dal modello contro l'avversario j-esimo.

Il parametro $K(t)$ rappresenta il cosiddetto *K-factor*, che determina la volatilità del rating, ovvero quanto questo debba reagire ai nuovi risultati. Nel modello di FiveThirtyEight, K non è una costante, bensì una funzione decrescente del numero di match giocati in carriera al tempo t, indicato con $m(t)$, e viene calcolato secondo la formula $K = 250/(m(t) + 5)^{0.4}$. Inoltre, in occasione dei tornei del Grande Slam, il *K-factor* viene moltiplicato per 1.1, assegnando di fatto ai match dei tornei *Major* un peso specifico superiore del 10% rispetto agli altri eventi del circuito.

Il motore predittivo dell'algoritmo, ovvero la probabilità attesa $\hat{\pi}_{ij}(t)$ che il giocatore i sconfigga il giocatore j nel match t, è calcolata tramite la funzione logistica applicata

alla differenza dei rispettivi rating:

$$\hat{\pi}_{ij}(t) = \frac{1}{1 + 10^{(E_j(t) - E_i(t))/400}} \tag{4}$$

Questa stima probabilistica tende asintoticamente a 1 all'aumentare del divario tra il rating del giocatore i e quello del suo avversario. Infine, per calibrare il sistema, a ogni tennista che fa il suo esordio nel circuito professionistico viene assegnato per convenzione un rating iniziale pari a $E(1) = 1500$.

3 Dati, Metodologia e Metriche di Valutazione

La transizione dai modelli teorici all'applicazione empirica richiede un'infrastruttura di dati solida e un rigoroso framework metodologico. In questo capitolo verrà illustrato il processo di raccolta e integrazione dei dati storici, le tecniche statistiche adottate per il trattamento dei dati mancanti e le metriche di valutazione probabilistica necessarie per quantificare le performance predittive degli algoritmi.

3.1 L'origine e l'integrazione dei dati storici

Per condurre un'analisi empirica esaustiva e replicare fedelmente le dinamiche del circuito professionistico, la fase di *data engineering* ha richiesto l'integrazione di due flussi informativi distinti.

La prima fonte è costituita dal database open-source curato da Jeff Sackmann (Tennis Abstract), universalmente riconosciuto nella letteratura accademica sportiva per la sua affidabilità. Questo archivio fornisce, per ogni match del circuito ATP, non solo l'esito finale e i dati anagrafici dei tennisti (età, altezza, mano dominante, ranking e punti ATP), ma anche statistiche disaggregate a livello di punto, quali il numero di ace, doppi falli, e la percentuale di punti vinti sulla prima e sulla seconda di servizio. Tali informazioni sono essenziali per la calibrazione dei modelli *point-based*.

La seconda fonte è rappresentata dall'archivio storico del portale *tennis-data.co.uk*, dal quale sono state estratte le quote pre-match offerte dalle agenzie di scommesse (bookmaker). L'unione dei due dataset è avvenuta mediante una chiave di integrazione basata sulla data dell'incontro e sui cognomi dei giocatori. Per evitare il fenomeno del *data leakage* (l'utilizzo di informazioni future per prevedere eventi passati), l'architettura del simulatore è stata progettata con un approccio *rolling window*: le probabilità per ogni singola giornata di incontri sono state stimate addestrando i modelli esclusivamente sui match disputati nei 365 giorni (o multipli) strettamente precedenti alla data in esame.

3.2 Trattamento e imputazione dei dati mancanti

A causa della disomogeneità connaturata al circuito tennistico (presenza di giocatori provenienti dalle qualificazioni, *wild card* o rientri da lunghi infortuni), l'assenza temporanea di alcuni dati storici rappresenta una problematica strutturale. L'eliminazione in blocco delle osservazioni incomplete avrebbe comportato una distorsione del campione, alterando la reale composizione dei tornei. Si è pertanto optato per l'imputazione dei valori mancanti.

Nello specifico, ai tennisti privi di una classifica ufficiale o situati in posizioni marginali è stato assegnato un valore di ranking convenzionale pari alla media dei giocatori

posizionati oltre la centesima posizione mondiale. Parallelamente, per i modelli basati sulle regressioni, a tutti gli atleti non accreditati di una testa di serie (*unseeded*) è stato assegnato il valore fittizio di 33, neutralizzando così le distorsioni generate dai valori nulli pur mantenendo intatta la distanza matematica dai giocatori di vertice.

Per i modelli *point-based*, nel caso in cui un giocatore non disponesse di un numero statisticamente significativo di match pregressi (fissato empiricamente a un minimo di 5 incontri), le sue probabilità di vittoria al servizio e in risposta sono state stimate utilizzando le medie aggregate della fascia di ranking di appartenenza (Top 100 o fuori dalla Top 100). Questo approccio garantisce che ogni previsione sia guidata dal livello atteso di abilità in assenza di dati individuali consolidati.

3.3 Metriche di valutazione probabilistica: Accuracy, Brier Score e Log-Loss

La previsione dell'esito di un incontro di tennis non si limita all'indicazione deterministica del vincitore, ma richiede l'espressione di un grado di incertezza sotto forma di probabilità continua. Per valutare la bontà dei modelli predittivi, sono state impiegate tre metriche fondamentali.

3.3.1 Accuratezza

L'accuratezza è la misura predittiva più intuitiva, definita come la percentuale di incontri in cui il modello ha indicato correttamente il vincitore. Operativamente, la vittoria viene assegnata al giocatore al quale il modello attribuisce una probabilità superiore al 50%.

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} I(\hat{y}_i = y_i)$$

dove I è la funzione indicatrice, $y_i \in \{0, 1\}$ rappresenta l'esito reale dell'incontro e $\hat{y}_i$ la previsione dicotomica del modello. Nonostante la sua immediata interpretabilità, l'accuratezza penalizza indistintamente ogni errore, ignorando il livello di confidenza associato alla previsione.

3.3.2 Brier Score

Per superare i limiti dell'accuratezza, si introduce il *Brier Score*, che misura l'errore quadratico medio delle previsioni probabilistiche.

$$Brier = \frac{1}{N}\sum_{i=1}^{N}(\hat{\pi}_i - y_i)^2$$

In questa equazione, $\hat{\pi}_i$ rappresenta la probabilità di vittoria predetta dal modello per il giocatore favorito, e y_i assume valore 1 se il giocatore vince e 0 in caso di sconfitta. Il

Brier Score assume valori compresi tra 0 (previsione perfetta) e 1 (errore massimo). A differenza dell'accuratezza, questa metrica premia i modelli ben calibrati, penalizzando in modo quadratico le previsioni che si discostano fortemente dall'esito reale.

3.3.3 Log-loss

La *Logarithmic loss* è una funzione di costo strettamente legata alla teoria dell'informazione e ampiamente utilizzata nel campo del machine learning e del betting sportivo. Per un campione di N match, la funzione è calcolata come segue:

$$LogLoss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{\pi}_i) + (1 - y_i)\log(1 - \hat{\pi}_i)]$$

La proprietà distintiva della Log-loss è l'applicazione di una penalità logaritmica. Questo significa che un modello subirà una sanzione gravissima qualora esprima un'altissima sicurezza su un esito che si rivela poi errato (ad esempio, assegnando una probabilità del 99% a un tennista che poi perde l'incontro). Essendo una metrica fondamentale in ambito finanziario, minimizzare la Log-Loss equivale a minimizzare l'overconfidence.

3.4 Il Bookmaker Consensus Model (BCM)

Per stabilire un termine di paragone rigoroso per i modelli, è fondamentale confrontarne le stime con quelle prodotte dal mercato reale. Come evidenziato dalla teoria dei mercati efficienti, le agenzie di scommesse incorporano nelle loro quote non solo le statistiche storiche, ma anche informazioni qualitative e asimmetriche (infortuni, motivazione, affaticamento, condizioni meteo).[13]

A tal fine, è stato implementato il *Bookmaker Consensus Model* (BCM). Per ogni match, le probabilità di base vengono estratte calcolando il reciproco della media delle quote offerte dai bookmaker. Tuttavia, la somma dei reciproci delle quote è sempre superiore a 1 a causa dell'aggio, ovvero il margine di profitto garantito dal banco. Per ottenere la reale aspettativa implicita del mercato, è necessario rimuovere matematicamente questo margine. Nel presente elaborato, le probabilità del BCM sono state ottenute attraverso una normalizzazione standard:

$$\pi_{BCM} = \frac{\frac{1}{q_v}}{\frac{1}{q_v} + \frac{1}{q_s}}$$

dove q_v e q_s rappresentano rispettivamente la quota aggregata offerta per il tennista vincente e quella offerta per il tennista sconfitto. Il BCM agirà nel presente studio come il *benchmark* supremo: il limite asintotico verso cui l'accuratezza dei modelli statistici puri mira a convergere.

4 Replica e Validazione dei Modelli Predittivi

Il fulcro di questo elaborato risiede nella replica e nell'estensione del fondamentale benchmark proposto da Kovalchik [1]. L'obiettivo di questo capitolo è testare le diverse famiglie di modelli probabilistici discusse nel Capitolo 2, valutandone le performance sul circuito ATP per la stagione 2014, e confrontandole con le probabilità implicite offerte dal mercato delle scommesse sportive.

4.1 Il benchmark di Kovalchik

Prima della pubblicazione dello studio di Stephanie Kovalchik, la letteratura in ambito *tennis analytics* risultava frammentata. Ogni nuovo modello predittivo veniva testato dagli autori su campioni di dati differenti, con il limite principale di essere testati su specifiche stagioni, diverse tra loro, rendendo impossibile un confronto oggettivo tra le varie metodologie di analisi.

Il merito del lavoro di Kovalchik è stato quello di uniformare il terreno di prova. L'autrice ha selezionato 12 modelli pubblicati in letteratura, suddivisi, come anticipato, tra modelli basati sulla regressione, modelli *point-based* e modelli basati sul confronto a coppie, e li ha testati in modo retrospettivo sugli incontri di singolare maschile del circuito ATP disputati nel 2014.

Per garantire la massima fedeltà metodologica, la replica condotta in questo elaborato ha adottato il medesimo meccanismo a finestra mobile (*rolling window*). Per ogni match disputato nel 2014, le probabilità di vittoria sono state stimate addestrando i modelli di regressione e le catene di Markov esclusivamente sui dati relativi ai 365 giorni precedenti la data dell'incontro. L'unica eccezione è stata fatta per i modelli che richiedono specificatamente una finestra temporale maggiore sulla quale essere addestrati, ad esempio il modello *ELO career wins*.

Questo approccio simula le esatte condizioni informative a disposizione di uno scommettitore o di un analista prima che i giocatori scendano in campo.

4.2 Il modello Bradley-Terry

Un'estensione originale introdotta in questo elaborato rispetto al benchmark di Kovalchik riguarda la modellazione *Paired Comparison*. Mentre i modelli Bradley-Terry classici stimano le abilità latenti dei tennisti basandosi unicamente sul risultato finale dell'incontro, il presente lavoro implementa l'analisi proposta da McHale e Morton [11].

Per comprendere l'inserimento di questa modellazione all'interno dell'elaborato, è necessario formalizzare matematicamente le differenze strutturali tra i due approcci.

4.2.1 Bradley-Terry applicato al match (Kovalchik)

Nei modelli Bradley-Terry testati da Kovalchik, l'unità di misura è il match. Siano α_i e α_j le abilità latenti di due giocatori i e j. La probabilità π_{ij} che il giocatore i vinca l'incontro contro j è definita come:

$$\pi_{ij} = P(Y_{ij} = 1) = \frac{\alpha_i}{\alpha_i + \alpha_j} \tag{5}$$

dove Y_{ij} è una variabile dicotomica che assume valore 1 in caso di vittoria del giocatore i e 0 altrimenti. La stima delle abilità si ottiene massimizzando la funzione di log-verosimiglianza su un set di N match storici. In uno dei due modelli analizzati nel proprio lavoro, considerando l'estensione dell'arco temporale ai due anni precedenti ai match, Kovalchik introduce una funzione di pesatura temporale a gradino (*step decay*): ai match disputati nei 12 mesi precedenti viene assegnato un peso $w = 1$, mentre ai match risalenti a 12-24 mesi prima viene assegnato un peso dimezzato $w = 0.5$.

4.2.2 Bradley-Terry applicato ai game (McHale & Morton)

Il modello proposto da McHale e Morton rivoluziona questa architettura classica proposta dalla letteratura spostando l'unità di osservazione dal match al singolo game. Assumendo che la probabilità di vincere un game sia anch'essa proporzionale al rapporto tra le abilità latenti dei giocatori, il contributo alla verosimiglianza di un singolo match in cui il giocatore i vince g_i game e il giocatore j vince g_j game può essere scritto come:

$$L(\alpha_i, \alpha_j) \propto \left(\frac{\alpha_i}{\alpha_i + \alpha_j}\right)^{g_i} \left(\frac{\alpha_j}{\alpha_i + \alpha_j}\right)^{g_j} = \frac{\alpha_i^{g_i} \alpha_j^{g_j}}{(\alpha_i + \alpha_j)^{g_i + g_j}} \tag{6}$$

Inoltre, invece di utilizzare un decadimento a gradini per valutare lo stato di forma recente, gli autori inseriscono una funzione di decadimento esponenziale continuo basato sull'emivita (*half-life*) controllata dal parametro ϵ, introducendo un parametro S_k per scalare l'importanza dei match storici giocati su una superficie diversa da quella del match da prevedere. Di conseguenza, per prevedere un match al tempo t su una specifica superficie S, la funzione di verosimiglianza su tutti i match passati (indice $k \in A_t$) viene estesa come segue:

$$L(\alpha_i(t,S); i = 1, \ldots, n) = \prod_{k \in A_t} \left[\frac{\alpha_i(t,S)^{g_i} \alpha_j(t,S)^{g_j}}{(\alpha_i(t,S) + \alpha_j(t,S))^{g_i + g_j}}\right]^{\exp(\epsilon(t - t_k))S_k} \tag{7}$$

Di fondamentale importanza è la definizione dell'emivita (ϵ) e del peso della superficie (S_k) che non vengono trattati come parametri stimati tramite massima verosimiglianza, bensì come iperparametri ottimizzati empiricamente tramite *grid search* per massimizzare le performance predittive (*out-of-sample*). Di seguito i valori ottimali di emivita e pesi di superficie per categoria di superficie, ottimizzati per le diverse misure di accuratezza predittiva stabiliti dagli autori [11].

Tabella 1: Valori ottimali di emivita e pesi di superficie per categoria di superficie, ottimizzati per le diverse misure di accuratezza predittiva.

Superficie	Misura	Emivita (giorni)	Peso Hardcourt	Peso Clay	Peso Grass
Hard/Carpet	m1	120	1	0.25	0.5
	m2	240	1	0.01	0.01
	m3	120	1	0.25	0.5
Clay	m1	120	0.25	1	0.01
	m2	240	0.01	1	0.01
	m3	120	0.01	1	0.01
Grass	m1	480	0.5	0.01	1
	m2	180	0.01	0.01	1
	m3	240	0.25	0.01	1

Questa modellazione offre tre vantaggi statistici sostanziali che giustificano la superiorità del modello di McHale e Morton nella nostra replica empirica:

1. **Maggiore sensibilità informativa:** Il margine di vittoria viene pienamente catturato dall'algoritmo. Sconfiggere un avversario con un netto 6-1 6-2 fornisce al modello un'informazione sul divario di forza ($\theta_i - \theta_j$) molto diversa rispetto a una vittoria lottata per 7-6 7-6, sfumatura che il modello di Kovalchik ignora trattandole entrambe come un semplice $Y_{ij} = 1$.

2. **Decadimento temporale continuo:** Anziché utilizzare uno *step decay* arbitrario, il modello applica una funzione di decadimento esponenziale continuo $w(t) = e^{-\lambda t}$, dove t rappresenta i giorni trascorsi dal match e λ è calibrato su un parametro di *half-life* (emivita) dipendente dalla superficie di gioco. Questo permette all'algoritmo di adattarsi in modo molto più fluido e reattivo alle dinamiche di forma degli atleti.

3. **Mitigazione della separazione perfetta:** Da un punto di vista puramente computazionale, computare i singoli game previene intrinsecamente i frequenti problemi di *Complete Separation*. Mentre è estremamente comune che un giocatore esordiente perda il 100% dei match disputati nel suo primo anno (mandando in divergenza la stima di α_i), è statisticamente quasi impossibile che perda il 100% dei *game* giocati, garantendo così una stima finita e calcolabile per α_i.

4.3 Complete Separation e Firth's Bias Reduction

Durante la fase di calibrazione in ambiente R dei modelli Bradley-Terry proposti da Kovalchik, è emersa una severa criticità statistica nota come *Complete Separation* (o separazione perfetta).

Nei modelli di regressione logistica e nel Bradley-Terry, la separazione perfetta si verifica quando un predittore discrimina perfettamente la variabile risposta. Nel contesto del tennis, ciò accade frequentemente per i giocatori che hanno disputato pochissimi incontri

in un determinato lasso di tempo (ad esempio, un esordiente che ha vinto l'unico match giocato, registrando un *win-rate* del 100%, o un qualificato che lo ha perso, registrando lo 0%). In questi casi, la stima di massima verosimiglianza diverge, producendo stime dei coefficienti tendenti a infinito. Questo si traduce in probabilità predette estremizzate (pari esattamente a 1 o 0), le quali, in caso di previsione errata, generano un valore di Log-Loss infinito, invalidando l'intera valutazione del modello.

Per ovviare a questa distorsione matematica, si è resa necessaria l'applicazione della *Firth's Bias Reduction* (verosimiglianza penalizzata) [14]. Il metodo di Firth introduce un termine di penalizzazione basato sulla matrice di informazione di Fisher direttamente nella funzione di verosimiglianza:

$$L^*(\beta) = L(\beta)|I(\beta)|^{1/2} \tag{8}$$

Questa correzione ammorbidisce le stime, garantendo la convergenza dell'algoritmo e riportando le probabilità predette su valori finiti ed empiricamente validi per il calcolo delle funzioni di costo.

4.4 Replica e validazione McHale & Morton

Il primo passo dell'analisi empirica consiste nella stima delle abilità latenti dei giocatori (λ_i) per confermare la solidità dell'infrastruttura matematica del modello Bradley-Terry. Nelle Tabelle 2 e 3 vengono presentati i risultati della replica a confronto con le misurazioni originali degli autori per l'anno 2008.

Tabella 2: Confronto "I 15 migliori giocatori alla fine del 2008 su tutte le superfici".

McHale & Morton				
Rank	Player	Model rating	Rating SE	ATP Ranking
1	Nadal R.	1.03	0.07	1
2	Federer R.	1.00	0.00	2
3	Djokovic N.	0.90	0.06	3
4	Murray A.	0.86	0.06	4
5	Davydenko N.	0.82	0.05	5
6	Roddick A.	0.81	0.05	8
7	Del Potro J.M.	0.81	0.06	9
8	Soderling R.	0.80	0.06	17
9	Hewitt L.	0.79	0.07	67
10	Nalbandian D.	0.78	0.06	11
11	Gasquet R.	0.76	0.06	25
12	Ferrer D.	0.76	0.05	12
13	Tsonga J.W.	0.76	0.06	6
14	Kiefer N.	0.74	0.06	38
15	Berdych T.	0.74	0.05	20

Replica				
Rank	Player	Model rating	Rating SE	ATP Ranking
1	Nadal R.	1.03	0.07	1
2	Federer R.	1.00	0.00	2
3	Djokovic N.	0.90	0.06	3
4	Murray A.	0.87	0.06	4
5	Davydenko N.	0.82	0.06	5
6	Roddick A.	0.81	0.06	8
7	Del Potro J.M.	0.80	0.06	9
8	Soderling R.	0.80	0.06	17
9	Hewitt L.	0.79	0.07	67
10	Nalbandian D.	0.78	0.06	11
11	Ferrer D.	0.76	0.06	12
12	Gasquet R.	0.76	0.06	25
13	Tsonga J.W.	0.75	0.06	6
14	Berdych T.	0.74	0.06	20
15	Ancic M.	0.73	0.06	36

Tabella 3: Confronto "I 10 migliori giocatori alla fine del 2008 basato esclusivamente sui risultati di cemento e terra rossa usando una funzione di decadimento con una *half-life* di 240 giorni".

McHale & Morton				
Rank	Player	Rating: hardcourt	Player	Rating: clay
1	Federer R.	1.00	Nadal R.	1.41
2	Djokovic N.	0.93	Federer R.	1.00
3	Nadal R.	0.90	Djokovic N.	0.91
4	Murray A.	0.89	Ferrer D.	0.80
5	Davydenko N.	0.86	Stepanek R.	0.79
6	Soderling R.	0.85	Hewitt L.	0.79
7	Nalbandian D.	0.84	Davydenko N.	0.78
8	Del Potro J.M.	0.84	Del Potro J.M.	0.77
9	Roddick A.	0.83	Gremelmayr D.	0.75
10	Tsonga J.W.	0.78	Nalbandian D.	0.74

Replica				
Rank	Player	Rating: hardcourt	Player	Rating: clay
1	Federer R.	1.00	Nadal R.	1.41
2	Djokovic N.	0.91	Federer R.	1.00
3	Murray A.	0.89	Djokovic N.	0.95
4	Nadal R.	0.88	Ferrer D.	0.80
5	Davydenko N.	0.86	Davydenko N.	0.78
6	Soderling R.	0.84	Del Potro J.M.	0.78
7	Nalbandian D.	0.83	Hewitt L.	0.77
8	Roddick A.	0.83	Stepanek R.	0.76
9	Del Potro J.M.	0.83	Gonzalez F.	0.74
10	Tsonga J.W.	0.76	Robredo T.	0.74

Come si evince dalle Tabelle 2 e 3, le stime delle abilità latenti replicate si sovrappongono in modo ottimale ai risultati del paper originale. Questa eccezionale precisione parametrica certifica l'assoluta correttezza dell'infrastruttura computazionale implementata, validando il motore matematico prima della sua applicazione esplorativa sul circuito ATP 2019-2024. Inoltre si mostra una divergenza tra il ranking ATP e l'abilità latente stimata dal modello: il fatto che i primi 5 giocatori coincidano perfettamente tra ATP Ranking e Model Rating conferma che al vertice la gerarchia ufficiale è un ottimo proxy della forza reale, tuttavia, scendendo di posizione, la presenza di giocatori nella "Top 15" come Lleyton Hewitt (Rank 9 nel modello, ma 67 nell'ATP nel 2008), suggerisce che il modello isola la capacità tecnica pura su una specifica superficie al netto della continuità stagionale richiesta dal ranking ufficiale, identificando un giocatore sottovalutato dal sistema di punti, potenzialmente per lunghi infortuni o scelte di calendario limitate.

Successivamente alla stima delle abilità latenti sulle singole superfici, l'analisi procede con la replica delle *"measures of predictive performance"*, così come definite dagli autori, allo scopo di valutare oggettivamente l'accuratezza probabilistica e il ritorno economico del modello applicato in un contesto di scommesse sportive:

- $m_1 = \frac{1}{N}\sum_{k\in B}\log(p_k^w)$: **Log-Loss**, rappresenta la media del logaritmo naturale della probabilità assegnata al giocatore che ha effettivamente vinto. Misura la penalizzazione per le previsioni errate formulate con elevata confidenza.
- $m_2 = \frac{1}{N}\sum_{k\in B} p_k^w$: **Accuracy Ponderata**, rappresenta la media aritmetica semplice delle probabilità assegnate ai vincitori effettivi.
- m_3: **Return on Investment a Quote Medie**, quantifica il profitto o la perdita percentuale simulando scommesse con le quote medie del mercato. La strategia prevede di puntare 1 unità esclusivamente in presenza di "Valore Atteso Positivo" (*Expected Value* > 0), ovvero quando $P_{\text{modello}} \times \text{Quota} > 1$.
- m_4: **Return on Investment a Quote Massime**, applica la medesima strategia di m_3, ottimizzando tuttavia il rendimento tramite la selezione della miglior quota (massima) disponibile sul mercato al momento dell'evento.

Tabella 4: Confronto "Misure della capacità predittive del modello e delle previsioni ricavate dalle quote dei bookmaker".

McHale & Morton

	Model				Bookmakers' odds			
	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
Hard/carpet	−0.620	0.598	−5.2%	1.6%	−0.590	0.585	−9.8%	−4.1%
Clay	−0.615	0.606	−5.2%	0.2%	−0.587	0.586	−10.5%	−5.1%
Grass	−0.575	0.618	−7.3%	−1.2%	−0.549	0.609	−13.0%	−7.6%

Replica

	Model				Bookmakers' odds			
	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
Hard/carpet	−0.710	0.560	−9.8%	−2.1%	−0.584	0.590	−9.9%	−4.2%
Clay	−0.737	0.567	−8.9%	−2.5%	−0.585	0.588	−10.1%	−4.3%
Grass	−0.745	0.560	−16.9%	−9.4%	−0.551	0.609	−12.8%	−7.3%

Per quanto concerne i parametri relativi all'efficienza dei *bookmaker*, i risultati ottenuti nell'elaborazione appaiono estremamente coerenti con le metriche originali dello studio, nonostante la presenza di parametri con leggeri scostamenti.

Tale discrepanza risulta giustificabile considerando la natura di una replica indipendente, nella quale non è possibile replicare in modo esatto le pratiche di filtraggio adottate

dagli autori (come la gestione tecnica dei walkover, l'inclusione di specifici tornei o il set esatto di *bookmaker* considerati nel calcolo delle medie).

Nonostante queste limitazioni sui dati di partenza, il risultato fondamentale emerso dalla Tabella 4 risiede nella conferma metodologica: si è riusciti a generare un modello realistico che, su superfici primarie quali cemento e terra battuta, è in grado di ridurre l'entità delle perdite rispetto allo scommettitore casuale, si osservi il confronto sul parametro m_4, validando l'intero framework.

4.5 Replica e validazione Kovalchik

Validato il motore matematico di McHale & Morton, l'obiettivo si è spostato sulla generazione di una valutazione comparativa in grado di mettere a confronto diverse famiglie di algoritmi predittivi.

I modelli sono stati valutati in base all'accuratezza e alla Log-Loss, e raffrontati sistematicamente con le aspettative espresse dal mercato.

Tabella 5: Riepilogo delle prestazioni predittive nei dati di validazione ATP del 2014, suddivise per tipologia.

Method	**Prediction accuracy**	**Calibration (95% CI)**	**Log-loss**
Regression-based			
Mean	65	0.96 (0.94, 0.99)	0.61
Basic probit	59	0.92 (0.88, 0.97)	0.63
Prize probit	68	0.98 (0.93, 1.03)	0.61
Probit plus	67	0.98 (0.93, 1.03)	0.60
Logistic	67	0.97 (0.92, 1.02)	0.60
Point-based			
Mean	64	0.91 (0.88, 0.93)	0.66
Basic IID	63	0.89 (0.85, 0.93)	0.67
Opponent adjusted	67	0.98 (0.93, 1.03)	0.63
Low level	64	0.87 (0.83, 0.91)	0.68
Common opponent	63	0.89 (0.84, 0.93)	0.66
Paired comparison			
Bradley-Terry 1	62	0.80 (0.76, 0.83)	0.67
Bradley-Terry 2	65	0.80 (0.76, 0.84)	0.65
FiveThirtyEight 1	67	0.98 (0.93, 1.03)	0.60
FiveThirtyEight 2	70	1.03 (0.98, 1.08)	0.59
BCM	72	0.98 (0.93, 1.03)	0.55

Tabella 6: Replica del riepilogo delle prestazioni predittive nei dati di validazione ATP del 2014, suddivise per tipologia.

Method	**Prediction accuracy**	**Calibration (95% CI)**	**Log-loss**
Regression-based			
Basic Probit	63	1.38 (1.35, 1.41)	0.64
Probit Plus	66	1.11 (1.08, 1.14)	0.65
Logistic 1	68	1.02 (0.99, 1.04)	0.60
Logistic 2	66	1.09 (1.06, 1.11)	0.61
Point-Based			
Basic IID	67	1.14 (1.11, 1.18)	0.63
Opponent adjusted	67	1.02 (0.99, 1.05)	0.61
Low level	63	1.07 (1.04, 1.10)	0.64
Common opponent	64	0.93 (0.90, 0.95)	0.65
Paired Comparison			
Bradley-Terry 1	64	0.94 (0.91, 0.96)	0.68
Bradley-Terry 2	66	0.96 (0.93, 0.98)	0.64
Bradley-Terry 3	66	0.211	0.62
FiveThirtyEight 1	68	0.96 (0.93, 0.98)	0.60
FiveThirtyEight 2	70	0.97 (0.95, 1.00)	0.59
BCM	72	1.03 (1.01, 1.06)	0.55

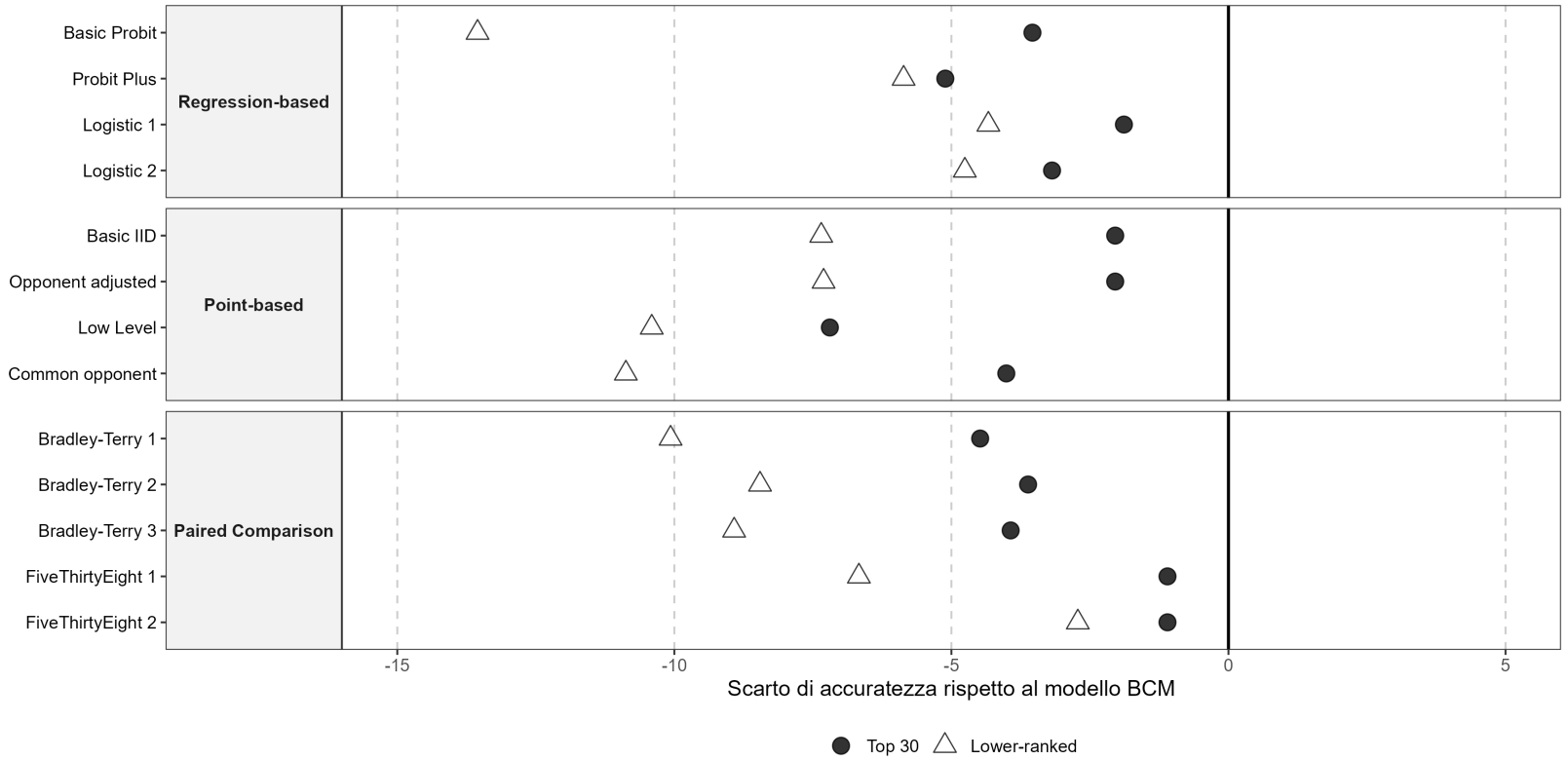

Figura 4.1: Differenza percentuale nell'accuratezza predittiva dei modelli replicati rispetto al modello di consenso dei bookmaker in base al ranking del giocatore meglio classificato nel match.

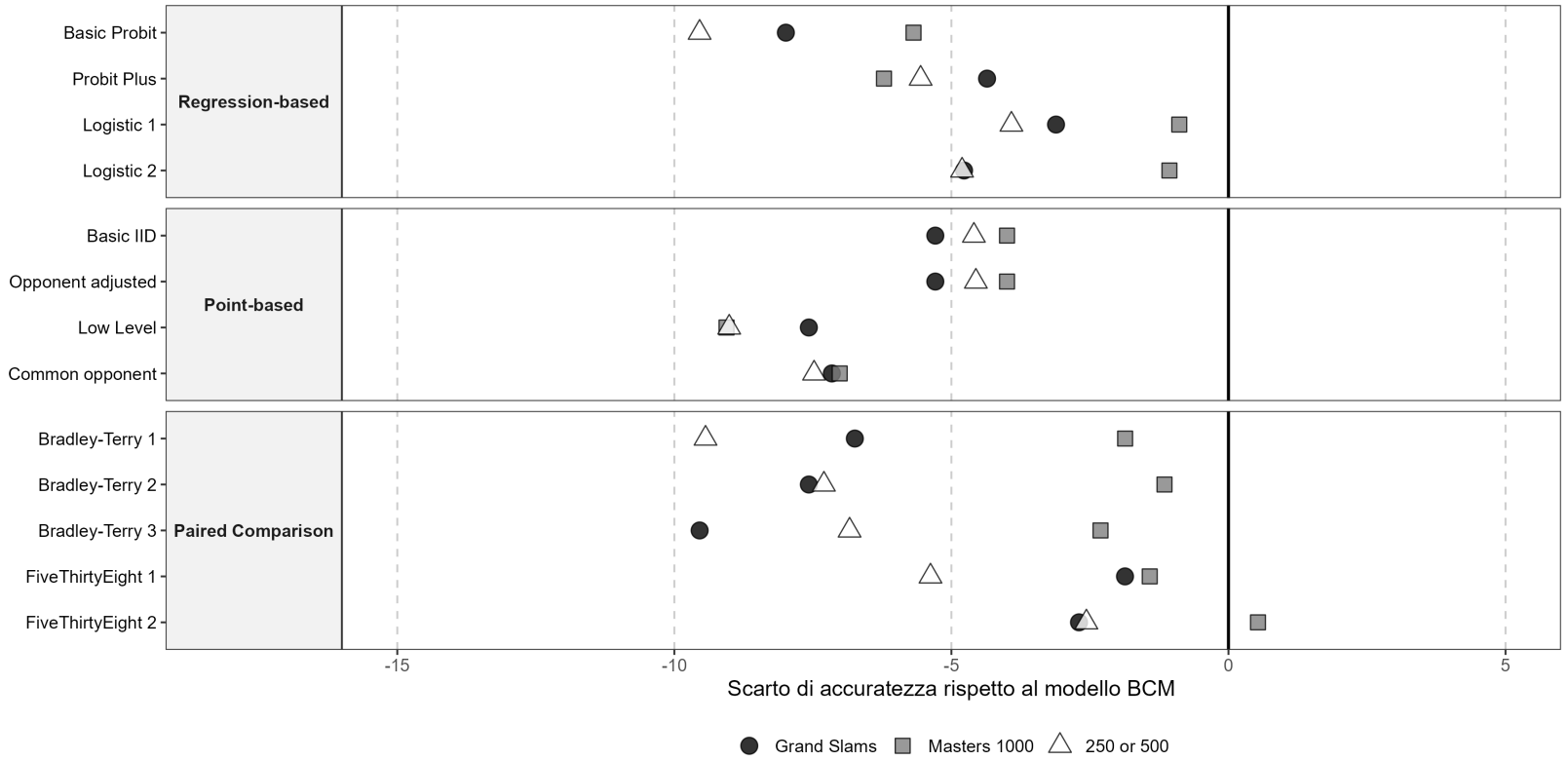

Figura 4.2: Differenza percentuale nell'accuratezza predittiva dei modelli replicati rispetto al modello di consenso dei bookmaker in base al livello del torneo, dalla fascia più alta (Grand Slam) a quella più bassa (Serie 250 e 500).

4.5.1 Analisi e discussione dei risultati

L'analisi strutturata dei risultati emersi dalla Tabella 6 permette di delineare le seguenti conclusioni chiave:

1. **Il benchmark insuperato del mercato:** conformemente a quanto originariamente evidenziato dalla letteratura [1], superare l'efficienza predittiva dei *bookmaker* rimane un'impresa estremamente ardua. Il modello BCM registra sistematicamente l'Accuratezza più elevata (72%) e la Log-Loss migliore (0.55). Questo risultato riafferma la natura fortemente efficiente del mercato del *betting* tennistico, capace di prezzare in tempo reale informazioni qualitative sfuggenti ai modelli puramente quantitativi.

2. **La superiorità dei Modelli Elo:** Tra i modelli puramente statistici, gli algoritmi basati sul sistema Elo emergono nettamente come i più performanti. La loro capacità intrinseca di auto-aggiornamento, che penalizza asimmetricamente le sconfitte contro avversari di ranking inferiore, garantisce un'accuracy del 70% (nella versione *Career*), distanziandosi di poco meno di due punti percentuali dal mercato e, sorprendentemente, battendolo se si analizza lo scaglione dei Master 1000 dell'accuratezza predittiva per livello di torneo.

3. **Il ruolo delle Regressioni Logistiche:** I modelli di regressione, nonostante l'apparente semplicità strutturale, si confermano strumenti predittivi di assoluta solidità. In particolare, il modello *Logistic 1*, fondato unicamente sul differenziale di ranking ATP, offre un rapporto costo-computazionale/beneficio eccellente registrano un 68% di accuratezza. Ciò dimostra che la gerarchia ATP, benché criticata, rappresenta un predittore naturale di notevole potenza esplicativa. Di contro, l'uso delle sole teste di serie (*Basic Probit*) produce risultati predittivi meno efficaci.

4. **La stabilità dei modelli Bradley-Terry:** L'inclusione del modello sviluppato da McHale & Morton (identificato in questa replica come *Bradley-Terry 3*) all'interno del framework di Kovalchik offre un riscontro importante. Basandosi sui game vinti/persi e implementando un decadimento temporale esteso a 3 anni, con un accuracy del 66.0% e una Log-loss pari a 0.62, primeggia nella categoria *Bradley-Terry*, confermando come la frammentazione del match nelle sue sotto componenti migliori le stime rispetto all'osservazione del solo esito finale.

5. **I limiti dei Modelli Point-Based:** I modelli Markoviani (*Point-Based*), pur vantando la maggiore eleganza teorica nel descrivere la struttura gerarchica tipica dei punteggi nel tennis, faticano a dominare la classifica generale. L'estrema varianza stocastica che caratterizza la vittoria del singolo punto al servizio rende l'aggregazione di tali probabilità instabile se utilizzata per estrapolare la vittoria del match, posizionando questi modelli in una fascia predittiva intermedia.

In conclusione, questa validazione empirica consolida un principio fondamentale della *tennis analytics*: nell'ottica della massimizzazione dell'accuratezza predittiva, l'aggiornamento dinamico dei punteggi (Elo) e le funzioni logit regresse sul ranking dominano statisticamente sulle inferenze punto per punto o sui modelli basati su abilità fisse.

5 Applicazione sul Circuito ATP 2019-2024

L'obiettivo finale di questo elaborato è quello di testare i modelli calibrati su dati storici all'interno del circuito contemporaneo, per testarne la validità nel presente. Questo confronto risulta un naturale sviluppo della ricerca, alla luce delle evoluzioni introdotte sia in ambito tennistico che in quello dell'analisi sportiva.

Il tennis ha vissuto negli ultimi anni profondi mutamenti: l'avvento dell'Hawk-Eye, come già anticipato, ha rivoluzionato l'analisi dei match grazie alla densità di dati che se ne possono estrapolare. Sempre dal lato dell'acquisizione dei dati, i sensori integrati nei manici delle racchette hanno permesso di raccogliere dati biomeccanici precisi durante gli scambi, analizzandone la potenza, il tipo di rotazione e l'angolo di impatto.

Il progresso tecnologico non si è limitato a cambiare esclusivamente l'approccio alle analisi ma ha permeato anche il gioco stesso: l'uso di grafite, carbonio e kevlar ha permesso di creare racchette più leggere, aerodinamiche e al tempo stesso estremamente rigide e stabili, facilitando la generazione di potenza da fondo campo, spostando il fulcro del gioco gradualmente dal *serve and volley* alla linea di fondo. Come principale conseguenza di questa transizione del gioco ne risulta l'aumento della lunghezza degli scambi, incentivato anche dall'omogeneizzazione delle superfici, tema di fondamentale importanza per comprendere le analisi che verranno svolte nei prossimi paragrafi. Se le nuove racchette hanno dato ai giocatori gli strumenti per colpire in topspin, sono stati i cambiamenti dei campi a creare l'ambiente perfetto per far prosperare questo stile di gioco ovunque.

Il ridimensionamento delle superfici si pone come risposta all'aumentare del disinteresse degli spettatori che stava colpendo il tennis di inizio millennio; il gioco era dominato dai grandi battitori e, nel caso di Wimbledon, gli scambi duravano in media da 1 a 3 colpi, con interi game risolti in ace o risposte sbagliate, con la conseguenza di rendere il gioco meno attrattivo per il pubblico. Il caso più dimostrativo è proprio quello relativo a Wimbledon: il celeberrimo Slam britannico, nonché l'unico Major giocato su erba, era noto per essere il torneo con la superficie tra le più veloci del circuito, per via di un prato scivoloso che si usurava rapidamente e faceva schizzare la pallina bassissima, rendendo difficoltoso rispondere da fondo campo. Nel 2001, per far fronte all'usura causata dai movimenti sempre più fisici dei giocatori e per promuovere il gioco da fondo campo, Wimbledon passò a un suolo molto più compatto e duro, che comportava un rimbalzo molto più alto delle palline e la perdita della componente di velocità orizzontale, sancendo il tramonto del *serve and volley*. Nel circuito ATP 2019-2024, dunque, a causa del livellamento delle superfici, non è più necessario cambiare radicalmente il proprio stile di gioco passando da una superficie all'altra. La costruzione del punto da fondo è diventata la strategia dominante su tutti i campi, portando così alla graduale scomparsa degli specialisti e assegnando più importanza alla tecnica e alle doti atletiche.

5.1 Il modello McHale & Morton contemporaneo

Nel paragrafo in questione si procederà ripetendo le medesime analisi condotte precedentemente nel paragrafo 4.4, con la principale differenza che l'input di riferimento non sono più i dati relativi alle stagioni dal 2000 al 2008 bensì quelli delle stagioni dal 2019 al 2024.

5.1.1 Stima delle abilità latenti

Tabella 7: "I 15 migliori giocatori alla fine del 2024 su tutte le superfici".

Rank	Player	Model Rating	SE	ATP Rank
1	Sinner J.	1.04	0.08	1
2	Djokovic N.	1.00	0.00	7
3	Alcaraz C.	0.98	0.08	3
4	Zverev A.	0.88	0.07	2
5	Medvedev D.	0.87	0.07	4
6	Draper J.	0.84	0.08	15
7	De Minaur A.	0.83	0.07	9
8	Fritz T.	0.82	0.07	5
9	Dimitrov G.	0.81	0.07	10
10	Paul T.	0.80	0.07	12
11	Rublev A.	0.79	0.07	8
12	Tsitsipas S.	0.79	0.07	11
13	Humbert U.	0.77	0.07	18
14	Nadal R.	0.76	0.11	261
15	Kyrgios N.	0.76	0.14	N/A

Tabella 8: "I 10 migliori giocatori alla fine del 2024 sul cemento".

Rank	Player	Model Rating	SE	ATP Rank
1	Sinner J.	1.05	0.10	1
2	Djokovic N.	1.00	0.00	7
3	Alcaraz C.	0.94	0.10	3
4	Medvedev D.	0.89	0.09	4
5	Zverev A.	0.87	0.08	2
6	Draper J.	0.84	0.09	15
7	Fritz T.	0.82	0.08	5
8	Dimitrov G.	0.80	0.08	10
9	De Minaur A.	0.80	0.08	9
10	Rublev A.	0.79	0.08	8

Tabella 9: "I 10 migliori giocatori alla fine del 2024 sull'erba".

Rank	Player	Model Rating	SE	ATP Rank
1	Alcaraz C.	1.02	0.16	3
2	Sinner J.	1.01	0.15	1
3	Djokovic N.	1.00	0.00	7
4	Zverev A.	0.90	0.14	2
5	Dimitrov G.	0.88	0.14	10
6	Medvedev D.	0.87	0.14	4
7	De Minaur A.	0.86	0.14	9
8	Nadal R.	0.85	0.32	261
9	Draper J.	0.85	0.14	15
10	Paul T.	0.83	0.13	12

Tabella 10: "I 10 migliori giocatori alla fine del 2024 sulla terra".

Rank	Player	Model Rating	SE	ATP Rank
1	Sinner J.	1.04	0.14	1
2	Alcaraz C.	1.00	0.14	3
3	Djokovic N.	1.00	0.00	7
4	Zverev A.	0.88	0.11	2
5	Tsitsipas S.	0.87	0.11	11
6	De Minaur A.	0.84	0.12	9
7	Fritz T.	0.83	0.11	5
8	Medvedev D.	0.82	0.11	4
9	Ruud C.	0.81	0.10	7
10	Draper J.	0.80	0.13	15

Osservando le Tabelle 7, 8, 9 e 10 relative al circuito ATP 2019-2024 e confrontandole con le Tabelle 2 e 3 del 2008, emergono due differenze strutturali fondamentali che fotografano l'evoluzione del tennis moderno. In primo luogo, si assiste a una marcata compressione delle abilità latenti. Nel 2008, specialisti come Nadal sulla terra rossa raggiungevano un rating di 1.41, evidenziando un divario incolmabile con il resto del circuito. Nel 2024, il leader su tutte le superfici (Jannik Sinner) registra punteggi compresi tra 1.01 e 1.05, indicando un vertice molto più denso, equilibrato e privo di dominatori assoluti. In secondo luogo, è evidente la scomparsa degli specialisti di superficie. Nel 2008, giocatori fuori dalla Top 50 ATP riuscivano a inserirsi nelle Top 10 per superficie, ad esempio Lleyton Hewitt al numero 67 se si considera il cemento. Oggi, le classifiche del modello per cemento, terra ed erba sono quasi perfettamente sovrapponibili alla Top 15 del ranking ufficiale ATP. I giocatori d'élite moderni sono tennisti "a tutto tondo", capaci di eccellere su ogni terreno grazie alla già citata omogeneizzazione dei campi, l'evoluzione dei materiali e la preparazione atletica. L'unica eccezione visibile nelle tabelle è rappresentata da Rafael Nadal (ATP 261, ma presente nella Top 15 del modello), tale anomalia non è un errore di calcolo, bensì una peculiarità matematica della funzione di decadimento esponenziale di McHale e Morton: a causa della prolungata inattività, Nadal non ha accumulato un numero di sconfitte sufficienti per abbattere il suo rating storico, portando l'algoritmo a valutare il giocatore sulle sue solide prestazioni passate.

5.1.2 Stima delle misure della capacità predittiva del modello

Tabella 11: Confronto "Misure della capacità predittiva del modello e delle previsioni ricavate dalle quote dei bookmaker".

McHale & Morton

	Model				Bookmakers' odds			
	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
Hard/carpet	−0.657	0.521	−6.9%	−0.8%	−0.584	0.595	−5.0%	−0.8%
Clay	−0.664	0.519	−5.8%	4.6%	−0.593	0.588	−3.8%	2.7%
Grass	−0.652	0.524	−10.7%	−4.8%	−0.569	0.605	−6.2%	−2.2%

I risultati esposti nella Tabella 11 confermano l'estrema difficoltà di generare profitto nel mercato moderno. Il modello Bradley-Terry basato sui game fatica a replicare le performance dei bookmaker: l'accuratezza (m_2) si ferma attorno al 52%, nettamente inferiore al 58-60 garantito dal mercato, con una Log-Loss (m_1) conseguentemente più penalizzata. Traducendo questi valori in ritorni economici, l'applicazione del modello porta a perdite sistematiche e strutturali (fino al -10.7% sull'erba a quote medie). L'unico rendimento teorico positivo si registra sulla terra battuta a quote massime ($m_4 = +4.6\%$). Tuttavia, alla luce dei dati complessivi, tale valore appare più come una fluttuazione statistica iso-

lata (o una specifica distorsione temporanea delle quote) piuttosto che la validazione di una strategia di scommessa profittevole e sostenibile nel lungo periodo.

5.1.3 Verifica dell'ipotesi di omogeneizzazione delle superfici

Uno degli interrogativi di ricerca più rilevanti riguarda l'impatto della progressiva omogeneizzazione delle superfici di gioco nel tennis contemporaneo.

Come già sottolineato, il modello Bradley-Terry attribuisce un'importanza cruciale ai pesi legati alla superficie: le probabilità di vittoria vengono stimate penalizzando fortemente i risultati pregressi ottenuti da un tennista su campi diversi da quello in cui si disputa il match corrente. Per verificare se tale assunto tecnico sia ancora valido o risulti obsoleto nel circuito post-2019, si è proceduto alla costruzione di una variante contrapposta al modello originale:

- **Bradley-Terry originale:** mantiene l'impostazione originale con decadimento temporale e *Half-life* a 240 giorni e pesi differenziati per superficie.
- **Bradley-Terry omogeneo:** mantiene la memoria lunga ma elimina la categorizzazione delle superfici, trattando i match su cemento, terra ed erba con pesi identici.

I risultati della simulazione condotta sul periodo 2019-2024 confermano l'ipotesi teorica di partenza. L'accuratezza del modello omogeneo si attesta al 61.0%, replicando esattamente il medesimo risultato del modello originale. Tale equivalenza si riscontra anche nell'analisi del Brier Score: arrotondando alla seconda cifra decimale, l'errore quadratico medio risulta stabilizzato sul valore di 0.23 per entrambe le varianti analizzate.

Al fine di conferire validità statistica a questa intuizione empirica, le distribuzioni dei risultati tra il modello originale e la variante sono state sottoposte al Test di McNemar per campioni appaiati, trattandosi di stime effettuate sul medesimo *set* di incontri. Il test ha restituito un *p-value* pari a 0.927730.

Essendo il *p-value* superiore alla soglia di significatività di $\alpha = 0.05$, si accetta l'ipotesi nulla di assenza di differenza tra i modelli. Dal punto di vista analitico, ciò conduce a una conclusione di notevole impatto: nel tennis contemporaneo, l'introduzione di rigidi parametri di penalizzazione basati sulla superficie non apporta alcun valore informativo aggiuntivo all'algoritmo, ma rischia di tradursi in *overfitting*. La forza intrinseca di un giocatore, valutata senza distinzione di suolo, è oggi un indicatore sufficiente per eguagliare i modelli storici.

5.2 I modelli di Kovalchik contemporanei

Nel paragrafo in questione si procederà ripetendo le medesime analisi condotte precedentemente nel paragrafo 4.5, con la principale differenza che l'input di riferimento non sono più i dati relativi alle stagioni dal 2000 al 2008 bensì quelli delle stagioni dal 2019 al 2024.

Tabella 12: Riepilogo delle prestazioni predittive nei dati di validazione ATP del 2024, suddivise per tipologia.

Method	**Prediction accuracy**	**Calibration (95% CI)**	**Log-loss**
Regression-based			
Basic Probit	60	1.37 (1.34, 1.40)	0.66
Probit Plus	62	1.06 (1.03, 1.09)	0.67
Logistic 1	64	0.99 (0.96, 1.02)	0.64
Logistic 2	64	1.02 (0.99, 1.05)	0.64
Point-Based			
Basic IID	62	1.10 (1.07, 1.13)	0.65
Opponent Adjusted	62	0.99 (0.96, 1.02)	0.64
Low Level	61	1.05 (1.02, 1.08)	0.66
Common Opponent	63	0.94 (0.91, 0.96)	0.67
Paired Comparison			
Bradley-Terry 1	60	0.90 (0.87, 0.92)	0.71
Bradley-Terry 2	62	0.94 (0.91, 0.96)	0.66
Bradley-Terry 3	61	1.13 (1.10, 1.16)	0.64
FiveThirtyEight 1	63	0.92 (0.90, 0.95)	0.64
FiveThirtyEight 2	64	0.92 (0.90, 0.95)	0.63
BCM	69	1.05 (1.02, 1.07)	0.59

Il confronto tra le performance predittive del 2014 (Tabella 6) e quelle del 2024 (Tabella 12) porta alla luce un fatto evidente: una degradazione dell'accuratezza per tutte le classi modellistiche. Nel 2014, i modelli dinamici come FiveThirtyEight 2 riuscivano a toccare il 70% di accuratezza, avvicinandosi al 72% del BCM. Nel 2024, il miglior modello statistico si ferma al 64%, e lo stesso limite superiore del mercato crolla al 69%.

Questa flessione non è imputabile a un malfunzionamento degli algoritmi ma a un radicale mutamento del panorama tennistico. Il 2014 era un'era dominata dai Big 3 (Federer, Nadal, Djokovic), caratterizzata da gerarchie inossidabili in cui i favoriti perdevano raramente contro i giocatori di bassa classifica. Il circuito contemporaneo del 2024 vanta invece una profondità atletica e tecnica senza precedenti: il divario fisico tra un Top 10 e un Top 50 si è assottigliato drasticamente, rendendo gli le vittorie degli sfavoriti molto più frequenti. Questo nuovo trend del tennis moderno ha abbassato di conseguenza il tetto massimo di prevedibilità dello sport.

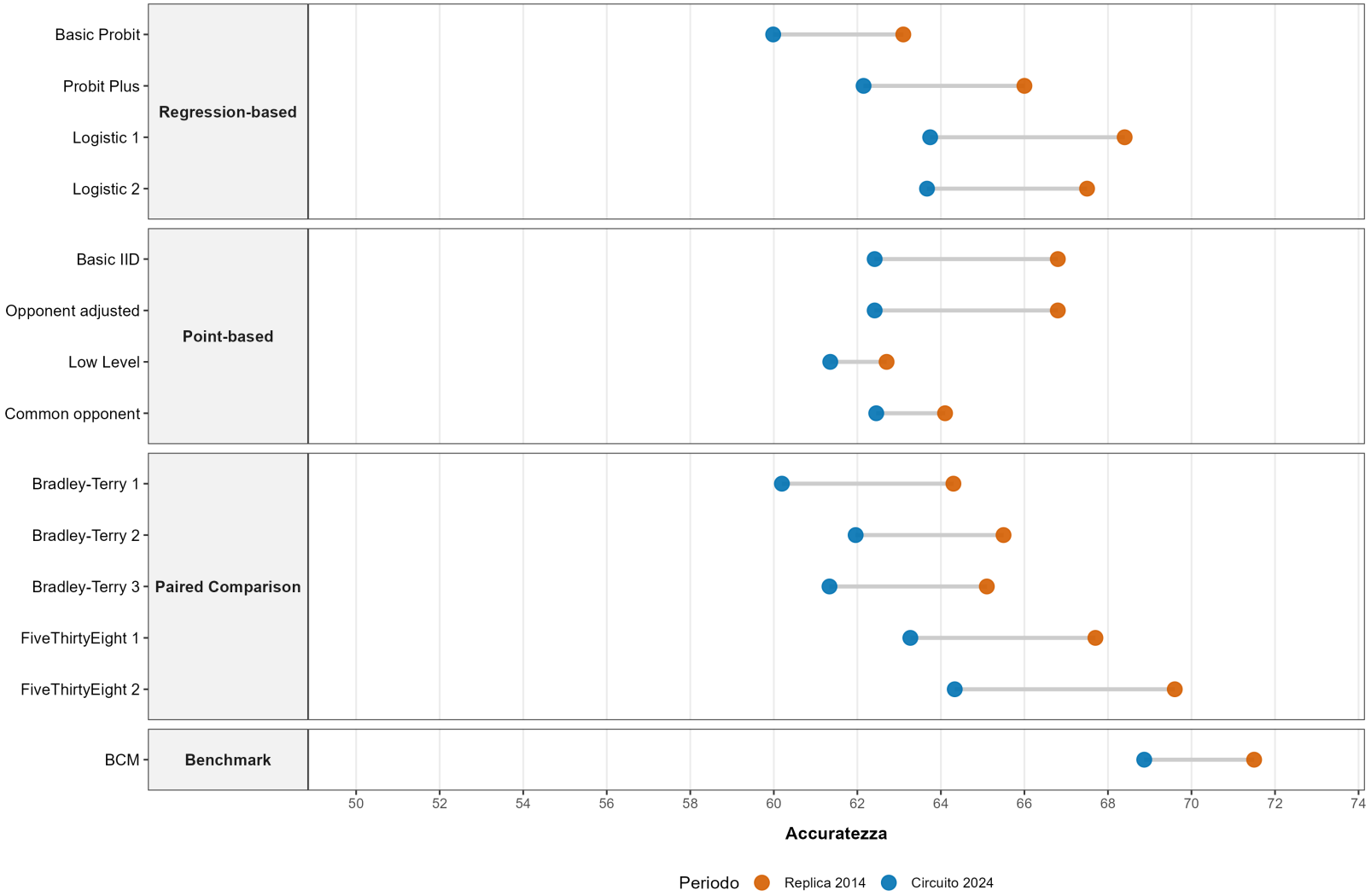

Figura 5.1: Confronto dell'evoluzione dell'accuratezza predittiva tra i modelli replicati e i modelli contemporanei.

5.2.1 Analisi di Significatività Statistica del Calo Predittivo

L'osservazione empirica dei risultati restituiti dai modelli applicati al circuito ATP 2019-2024 evidenzia una flessione generalizzata dell'accuratezza predittiva rispetto al benchmark storico del 2014. Tuttavia, in ambito inferenziale, l'osservazione di un calo percentuale non è sufficiente per trarre conclusioni strutturali, potendo tale flessione essere imputabile a una mera fluttuazione campionaria. Si è reso pertanto necessario verificare se il degrado delle performance predittive fosse *statisticamente significativo*.

Per validare tale ipotesi, è stato implementato un **Test Z per la differenza tra due proporzioni** a una coda. Per ciascun modello, si è confrontata la proporzione di successi ottenuta sul set di validazione del 2014 (p_{2014}) con quella ottenuta sul circuito del 2024 (p_{2024}). Al fine di mantenere una rigorosa coerenza matematica con la metrica di *Accuracy* precedentemente definita (Tabella 12), nei casi in cui il modello ha restituito una probabilità di vittoria esattamente pari a 0.5 (situazione di massima incertezza), si è assegnato convenzionalmente un valore di 0.5 successi.

Le ipotesi del test sono state formulate come segue:

- **Ipotesi Nulla (H_0):** $p_{2014} \leq p_{2024}$. Il calo osservato è frutto del caso; il circuito ATP 2019-2024 è prevedibile tanto quanto (o più) di quello storico.
- **Ipotesi Alternativa (H_1):** $p_{2014} > p_{2024}$. Il calo di accuratezza è reale e riflette un cambiamento strutturale nell'imprevedibilità del circuito.

Il livello di significatività (α) è stato fissato al 5% (0.05). I risultati del test sono riassunti nella Tabella 13.

Tabella 13: Risultati del Test Z a una coda per la differenza delle accuratezze dei modelli di replica e di quelli contemporanei.

Modello	**N. Match 2014**	**Acc. 2014**	**N. Match 2024**	**Acc. 2024**	**P-Value**
Regression-based					
Basic Probit	2449	63.1%	2634	60.0%	0.0115*
Probit Plus	2449	66.0%	2634	62.1%	0.0024*
Logistic 1	2449	68.4%	2634	63.7%	0.0003*
Logistic 2	2449	67.5%	2634	63.7%	0.0020*
Point-based					
Basic IID	2449	66.9%	2634	62.4%	0.0005*
Opponent adjusted	2449	66.9%	2634	62.4%	0.0005*
Low Level	2449	62.7%	2634	61.4%	0.1717
Common opponent	2449	64.2%	2634	62.5%	0.1048
Paired Comparison					
Bradley-Terry 1	2449	64.3%	2634	60.2%	0.0016*
Bradley-Terry 2	2449	65.5%	2634	62.0%	0.0048*
Bradley-Terry 3	2449	65.1%	2634	61.3%	0.0029*
FiveThirtyEight 1	2449	67.7%	2634	63.3%	0.0005*
FiveThirtyEight 2	2449	69.6%	2634	64.3%	<0.0001*
Benchmark					
Bookmakers (BCM)	2449	71.5%	2634	68.9%	0.0236*

** p-value < 0.05. Rifiuto dell'Ipotesi Nulla (H_0). Il calo di accuratezza è statisticamente significativo.*

5.2.2 Discussione dei Risultati

L'analisi inferenziale fornisce un supporto matematico inequivocabile alle dinamiche osservate graficamente, evidenziando tre fenomeni chiave:

1. **Il peggioramento dei modelli dominanti:** I modelli che nel 2014 eccellevano nell'estrarre segnale dai dati storici mostrano i cali più drastici e significativi. In particolare, il modello dinamico *FiveThirtyEight 2 (Elo Career)* subisce un crollo del 5.3% ($p < 0.0001$), mentre il modello logistico basato sui ranking (*Logistic 1*) perde il 4.7% ($p = 0.0003$). Questa discesa certifica che metriche storicamente infallibili, come il divario nel ranking o l'accumulo di punti Elo, hanno perso gran parte del loro potere discriminante nel circuito odierno. La varianza delle abilità latenti (il parametro α nei modelli Bradley-Terry) si è assottigliata: i giocatori di media classifica sono oggi atleticamente e tecnicamente molto più vicini alla Top 10 rispetto a un decennio fa, portando i modelli a generare previsioni che si schiacciano verso il limite teorico del 50%.

2. **L'aumento dell'imprevedibilità del mercato:** Il risultato più rilevante di questa analisi risiede nel p-value associato al *Bookmaker Consensus Model* ($p = 0.0236$). Il mercato delle quote prezza l'inclusione di tutte le informazioni disponibili (quantitative e qualitative, come infortuni, motivazioni e condizioni meteorologiche), rappresentando di fatto il tetto massimo di prevedibilità teorica dell'evento sportivo. Il fatto che il BCM subisca una flessione significativa (dal 71.5% al 68.9%) dimostra che l'imprevedibilità intrinseca del sistema "Tennis ATP" è oggettivamente aumentata. Non sono gli algoritmi matematici ad aver perso efficacia, ma è circuito post-Big 3 a essere divenuto strutturalmente più equilibrato, abbassando il limite superiore dell'informazione estraibile.

3. **Le eccezioni dei modelli Point-Based:** Gli unici modelli a non rigettare l'ipotesi nulla sono stati il *Low Level* ($p = 0.1717$) e il *Common opponent* ($p = 0.1048$). Questo non indica una superiorità di tali algoritmi nel 2024, bensì deriva dal fatto che questi modelli partivano già da una *baseline* prestazionale mediocre nel 2014. Il loro divario tra i due decenni risulta contenuto semplicemente perché, basandosi su assunzioni teoriche deboli per l'intero match (come l'ipotesi Markoviana di indipendenza dei punti i.i.d.), faticavano a estrarre valore predittivo in entrambe le ere tennistiche analizzate.

In conclusione, il test di significatività convalida l'ipotesi che la degradazione delle performance predittive nel circuito maschile contemporaneo sia un fenomeno non casuale. La riduzione del gap tecnico-atletico ha reso inefficaci i differenziali di punteggio rigidi, giustificando la necessità teorica, esplorata in precedenza, di adottare architetture *ensemble* per compensare la perdita di informazione dei singoli regressori.

5.3 Il Modello Ensemble

Alla luce della volatilità predittiva emersa nell'analisi comparata tra il 2014 e il 2024 (Tabella 13), è risultato evidente come l'affidamento a un singolo algoritmo sia divenuto un approccio intrinsecamente rischioso per la previsione degli esiti nel tennis contemporaneo. L'omogeneizzazione delle superfici e il conseguente appiattimento delle differenze tecniche hanno esposto i limiti specifici di ciascuna classe modellistica.

Per ovviare a tale vulnerabilità, si è optato per la costruzione di un Modello *ensemble*. L'impianto teorico di tale metodologia si basa sulla dimostrazione matematica che l'aggregazione lineare di più previsioni individuali, se fondate su insiemi di informazioni indipendenti, produce un errore quadratico medio e una varianza sistematicamente inferiori, o quantomeno uguali, rispetto alla migliore delle previsioni prese singolarmente [15]. Nel contesto specifico dello *sports forecasting*, l'approccio *ensemble* segue il processo decisionale dei bookmaker, i quali non operano basandosi su un singolo indicatore, bensì su un aggregato di metriche eterogenee [16].

5.3.1 Architettura del Modello

Al fine di massimizzare l'indipendenza dell'informazione elaborata, riducendo la multicollinearità tra i predittori, la strategia aggregativa prescelta è consistita in una media aritmetica semplice di tre algoritmi con specializzazioni divergenti:

1. **FiveThirtyEight 2:** valuta la forza intrinseca di lungo periodo e l'inerzia globale dell'atleta. Si caratterizza per un'elevata stabilità e una lenta reattività ai trend di breve termine.

2. **Bradley-Terry 3:** analizza le probabilità di vittoria del singolo game e di pesa le performance in funzione della superficie di gioco. Risulta altamente reattivo e specifico rispetto alle condizioni ambientali.

3. **Logistic 1:** una regressione logistica fondata esclusivamente sul ranking ATP, impiegata per catturare le gerarchie ufficiali del circuito.

L'assunto operativo è definito dalla seguente combinazione lineare:

$$P_{Ensemble} = \frac{P_{Elo} + P_{BT3} + P_{Logit}}{3} \tag{9}$$

Tale architettura obbedisce a un principio di compensazione statistica: l'errore generato dalla rigidità di un modello, ad esempio la sovrastima di un ex-campione in declino da parte dell'Elo, viene mitigato e corretto dall'intervento algoritmico degli altri, ad esempio il modello Bradley-Terry, maggiormente ancorato alla performance contingente.

5.3.2 Discussione dei Risultati e Limiti Strutturali

L'applicazione del modello ensemble sul dataset esclusivo del 2024 ha confermato gli assunti teorici della *forecast combination*. Sebbene l'accuratezza non sia riuscita a sfondare il tetto massimo dettato dal *Bookmaker Consensus Model*, attestandosi su valori in linea con i migliori modelli singoli, il super-modello ha manifestato un marcato incremento della robustezza statistica.

Come si evince dai dati, l'ensemble funge da filtro, ammortizzando le fluttuazioni estreme e garantendo una calibrazione significativamente più solida. Questo bilanciamento si riflette in metriche di penalizzazione, come la Log-loss, ottimizzate, dimostrando che l'aggregazione predittiva rappresenta la metodologia più adeguata per arginare l'incremento sistemico della varianza emerso nel tennis moderno.

Tuttavia, è necessario riconoscere i limiti strutturali che permangono in tale architettura, condivisi trasversalmente dall'intera modellistica analizzata. In primo luogo, l'assenza di variabili latenti biomeccaniche impedisce all'algoritmo di valutare lo stato di affaticamento o la suscettibilità alla pressione agonistica. In secondo luogo, persiste una distorsione nei modelli di base, più precisamente nei modelli *point-based*, derivante dall'assunzione che i punti disputati nel tennis siano I.I.D., un assioma noto in letteratura per essere non totalmente accurato a causa degli effetti del *momentum*.

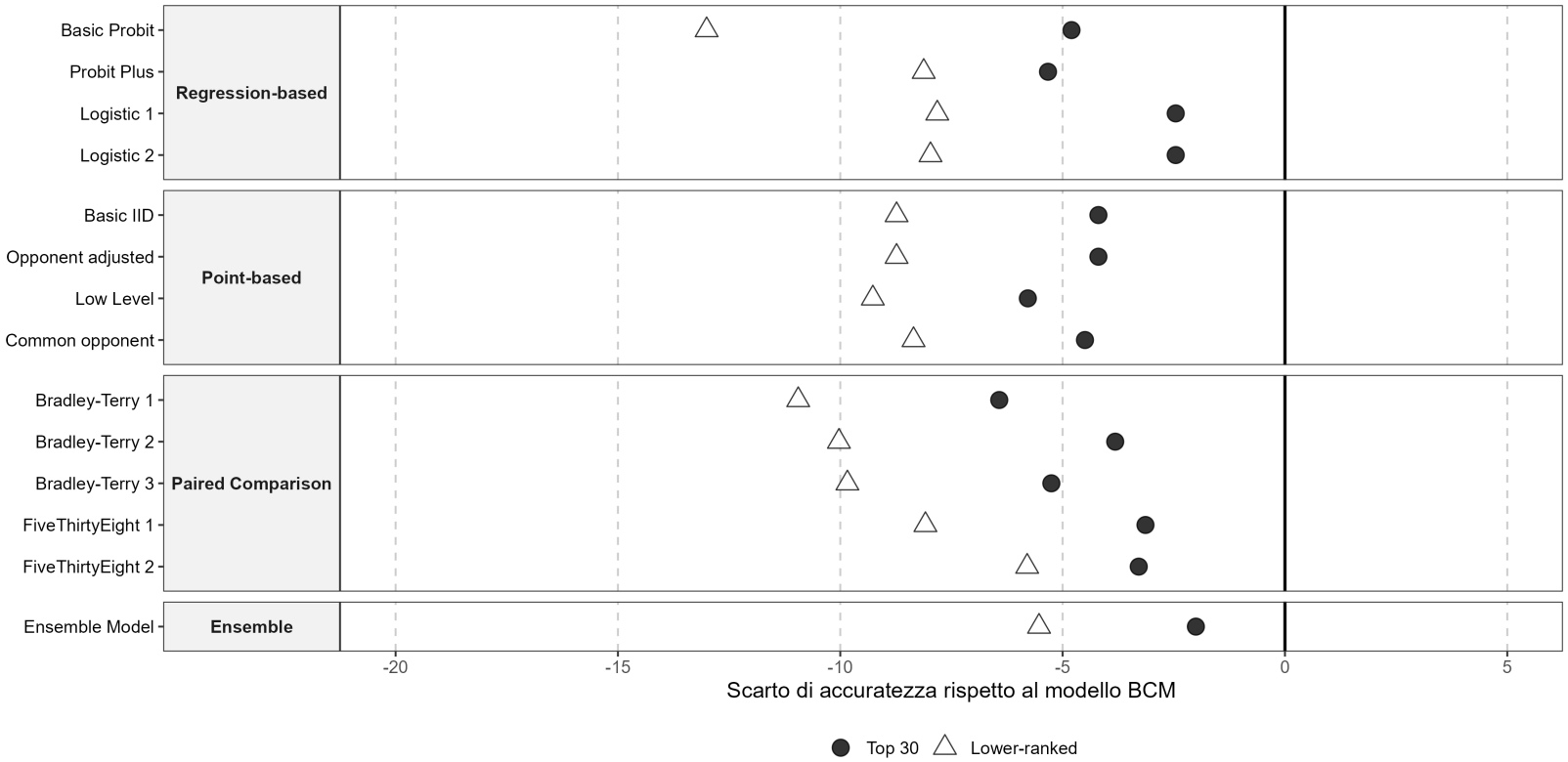

Figura 5.2: Differenza percentuale nell'accuratezza predittiva dei modelli contemporanei rispetto al modello di consenso dei bookmaker in base al ranking del giocatore meglio classificato nel match.

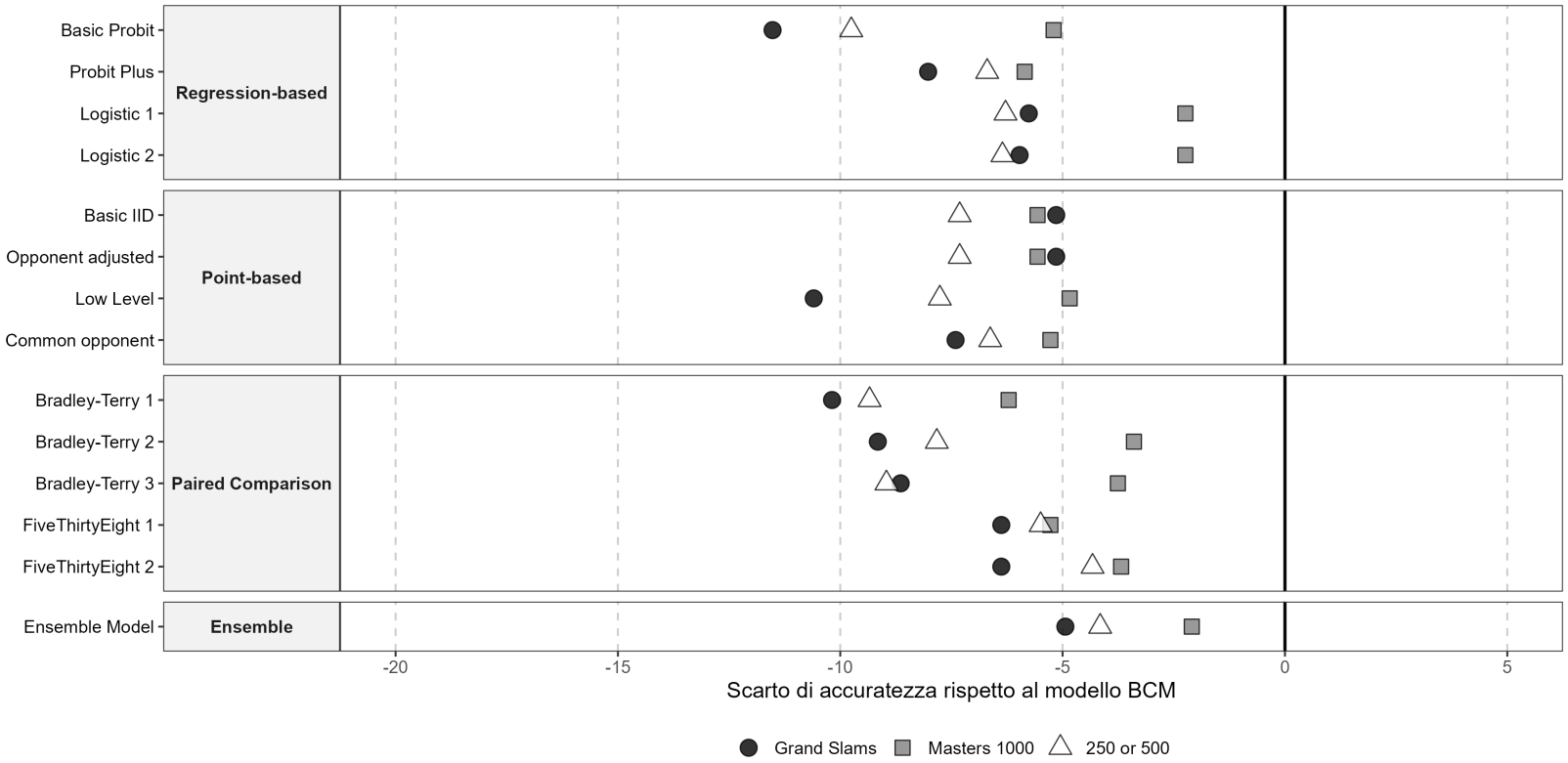

Figura 5.3: Differenza percentuale nell'accuratezza predittiva dei modelli contemporanei rispetto al modello di consenso dei bookmaker in base al livello del torneo, dalla fascia più alta (Grand Slam) a quella più bassa (Serie 250 e 500).

I grafici relativi allo scarto di accuratezza rispetto al modello di riferimento del mercato offrono preziose chiavi di lettura sui limiti dei modelli quantitativi rispetto alla conoscenza qualitativa dei bookmaker.

Analizzando le performance in base alle fasce di classifica (figura 5.2) emerge un pattern sistematico: tutti i modelli statistici riescono ad avvicinarsi maggiormente all'accuratezza del mercato quando prevedono match che coinvolgono giocatori in Top 30. Questo fenomeno, come è intuitivo pensare, è caratterizzato da una maggiore costanza di rendimento e da un volume di dati storici molto più ampio per i top player. Al contrario, quando in campo scendono giocatori fuori dai radar principali, il divario con i bookmaker si allarga. I tennisti di bassa classifica sono per natura più volatili, soggetti a rapidi cambi di forma o frequenti passaggi tra il circuito Challenger e quello ATP. In questo territorio ad alta varianza, le informazioni asimmetriche a disposizione dei bookmaker permettono al mercato di prezzare le quote in modo molto più efficiente di quanto possa fare un algoritmo basato sui risultati passati.

L'analisi per livello del torneo (figura 5.3) rivela una dinamica controintuitiva ma affascinante: i modelli matematici ottengono le loro migliori performance relative nei tornei Masters 1000, mentre faticano nettamente di più nei tornei del Grande Slam e nei tornei minori ATP 250 o 500. Questa stratificazione può essere interpretata analizzando la combinazione tra motivazione e formato di gioco:

1. **Masters 1000**: sono tornei a partecipazione obbligatoria per i top player, dove la motivazione è sempre massima, ma si disputano al meglio dei 3 set. Essendo il formato standard in cui si gioca la stragrande maggioranza dei match stagionali, i dati storici alimentano i modelli in modo coerente e pulito, permettendo alla statistica di rivaleggiare con i bookmaker.

2. **Grand Slams**: la motivazione è al vertice assoluto ma, giocandosi al meglio dei 5 set, questo formato introduce variabili fisiche, atletiche e psicologi che i modelli tradizionali faticano a ponderare correttamente, essendo addestrati prevalentemente su match al meglio dei 3 set. I bookmaker, invece, interpretano perfettamente questo fattore, staccando nettamente gli algoritmi in termini di accuratezza.

3. **ATP 250/500**:in questi tornei lo scarto torna ad ampliarsi a causa delle dinamiche motivazionali. I top player spesso utilizzano questi eventi come preparazione per gli Slam, giocando al di sotto dei propri standard o ritirandosi precauzionalmente ai primi sintomi di affaticamento. Il bookmaker, consapevole del contesto, adeguerà le quote incamerando un vantaggio predittivo che il modello non può replicare.

6 Conclusione e sviluppi futuri

Il presente elaborato si è posto l'obiettivo di sistematizzare, replicare e testare l'efficacia dei principali modelli predittivi statistici applicati al tennis, tracciando un ponte analitico tra il benchmark storico del 2014 e il panorama tennistico contemporaneo del 2024. L'analisi empirica condotta ha permesso non solo di validare infrastrutture matematiche note in letteratura, come i modelli di regressione e le catene di Markov, ma anche di adattare approcci complessi come il Bradley-Terry a livello di game implementato da McHale & Morton.

Il risultato fondamentale emerso dallo studio è duplice. In primo luogo, si è confermato che superare l'efficienza informativa del Bookmaker Consensus Model (BCM) attraverso l'esclusivo ausilio di dati retrospettivi e macro-statistiche (ranking, punti vinti, esiti dei game) rimane una sfida asintotica. Le asimmetrie informative a disposizione del mercato, infortuni, motivazioni, condizioni meteorologiche, garantiscono al banco un vantaggio sistematico impossibile da catturare per i modelli basati unicamente su serie storiche. In secondo luogo, il confronto statistico inferenziale tra le stagioni 2014 e 2024 ha certificato un radicale e significativo calo della prevedibilità generale del circuito ATP. Come dimostrato matematicamente, la combinazione tra omogeneizzazione delle superfici di gioco, standardizzazione dell'equipaggiamento e innalzamento della preparazione atletica ha causato una compressione delle abilità latenti (α). Il tennis odierno si configura, pertanto, come un ecosistema altamente denso e sfaccettato, in cui la scomparsa degli specialisti di superficie e l'assottigliamento del gap tra top player e giocatori di media classifica hanno reso obsoleti i vecchi metodi di previsione.

Per mitigare l'aumentata varianza del circuito, l'introduzione di un Modello Ensemble si è rivelata una scelta metodologica necessaria. L'aggregazione lineare di modelli complementari ha dimostrato che la robustezza predittiva, intesa come riduzione della Log-loss e ammortizzazione delle deviazioni estreme, rappresenta la frontiera ottimale per operare in un contesto ad alta incertezza.

Tuttavia, l'apparente raggiungimento di un limite per l'accuratezza dei modelli statistici non segna la fine dell'analisi quantitativa nel tennis, bensì ne definisce un momento di transizione. I limiti strutturali emersi in questo lavoro indicano chiaramente le direttrici per gli sviluppi futuri della Tennis Analytics.

La disponibilità sempre più massiccia di dati ottici e spaziali derivanti da sistemi di tracciamento come l'Hawk-Eye, e l'introduzione di metriche biomeccaniche catturate da wearable e racchette intelligenti, richiedono un cambio di paradigma. Le future architetture predittive dovranno abbandonare la logica riassuntiva del tabellino a fine match per abbracciare l'analisi dei dati posizionali (spatio-temporal data). L'integrazione di algoritmi di Machine Learning e Deep Learning – come le Reti Neurali Ricorrenti (RNN) o i modelli Transformer – permetterà di superare l'assunzione di indipendenza dei punti

(I.I.D.) che ha storicamente limitato i modelli markoviani. Tramite questi nuovi approcci sarà possibile quantificare dinamiche complesse e intangibili come l'inerzia psicologica (momentum), la gestione delle geometrie del campo e l'affaticamento muscolare durante la singola costruzione tattica dello scambio.

In sintesi, se i modelli statistici tradizionali analizzati in questo elaborato hanno esaurito la loro capacità di estrarre "nuovo segnale" dalle statistiche aggregate, essi restano le fondamenta matematiche imprescindibili su cui andranno innescati i futuri modelli di intelligenza artificiale, segnando il definitivo passaggio del tennis da uno sport letto tramite il risultato, a uno sport decodificato attraverso il puro movimento.

Bibliografia

[1] Kovalchik, Stephanie Ann. "Searching for the GOAT of tennis win prediction". In: *Journal of Quantitative Analysis in Sports* 12.3 (2016), pp. 127–138.

[2] Leitner, Christoph, Zeileis, Achim e Hornik, Kurt. *Is Federer Stronger in a Tournament without Nadal? An Evaluation of Odds and Seedings for Wimbledon 2009*. Rapp. tecn. 94. Research Report Series / Department of Statistics e Mathematics, 2009.

[3] Boulier, Bryan L. e Stekler, Herman O. "Are sports seedings good predictors?: an evaluation". In: *International Journal of Forecasting* 15.1 (1999), pp. 83–91.

[4] Klaassen, Franc J.G.M. e Magnus, Jan R. "Forecasting the winner of a tennis match". In: *European Journal of Operational Research* 148.2 (2003), pp. 257–267.

[5] Corral, Julio del e Prieto-Rodriguez, Juan. "The determinants of winning tennis matches: an investigation of the ATP tour". In: *Journal of Sports Economics* 11.5 (2010), pp. 501–527.

[6] Klaassen, Franc J.G.M. e Magnus, Jan R. "Are points in tennis independent and identically distributed? Evidence from Wimbledon". In: *Journal of the American Statistical Association* 96.454 (2001), pp. 500–509.

[7] Newton, Paul K. e Keller, Joseph B. "Probability of winning at tennis I. Theory and data". In: *Studies in Applied Mathematics* 114.3 (2005), pp. 241–269.

[8] Barnett, Tristan e Clarke, Stephen R. "Combining player statistics to predict outcomes of tennis matches". In: *IMA Journal of Management Mathematics* 16.2 (2005), pp. 113–120.

[9] Spanias, Demetris e Knottenbelt, William J. "Predicting the outcomes of tennis matches using a low-level point model". In: *IMA Journal of Management Mathematics* 24.3 (2012), pp. 311–320.

[10] Knottenbelt, William J., Spanias, Demetris e Madurska, Agnieszka M. "A common-opponent stochastic model for predicting tennis match results". In: *Computers & Mathematics with Applications* 64.12 (2012), pp. 3820–3827.

[11] McHale, Ian e Morton, Alex. "A Bradley-Terry type model for forecasting tennis match results". In: *International Journal of Forecasting* 27.2 (2011), pp. 619–630.

[12] Fischer-Baum, Reuben, Flowers, Andrew e Bialik, Carl. *How We Are Forecasting The 2015 US Open*. FiveThirtyEight. 2015. URL: `https://fivethirtyeight.com/features/how-we-are-forecasting-the-2015-us-open/` (visitato il giorno 20/05/2024).

[13] Forrest, David, Goddard, John e Jones, Robert. “Odds-setters as forecasters: The statistical efficiency of football betting odds”. In: *International Journal of Forecasting* 21.3 (2005), pp. 509–527.

[14] Firth, David. “Bias reduction of maximum likelihood estimates”. In: *Biometrika* 80.1 (1993), pp. 27–38.

[15] Bates, John M e Granger, Clive WJ. “The combination of forecasts”. In: *Journal of the Operational Research Society* 20.4 (1969), pp. 451–468.

[16] Constantinou, Anthony C e Fenton, Norman E. “Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries”. In: *Journal of Quantitative Analysis in Sports* 9.1 (2013), pp. 37–50.