

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE IN SCIENZE STATISTICHE

MODELLO BAYESIANO DINAMICO A RISPOSTA
ORDINALE PER L'ANALISI DEI RISULTATI CALCISTICI

Relatore:

Ch.mo Prof.
Nicola Sartori

Laureando:

Davide Bortoletto
Matricola: 2146880

Correlatore:

Ch.mo Dott.
Lorenzo Schiavon

Indice

Introduzione	1
1 Modelli statistici per l'analisi dei risultati calcistici	3
1.1 L'approccio <i>goal-based</i>	3
1.2 L'approccio <i>result-based</i>	7
1.3 Motivazione e proposta metodologica	8
2 Modellazione	11
2.1 Specificazione del modello	11
2.2 Distribuzioni a priori dei parametri	15
2.3 Inferenza bayesiana tramite metodi Monte Carlo	19
2.4 Verosimiglianza, distribuzione a posteriori e <i>full conditionals</i>	25
3 Applicazione	43
3.1 Dati e struttura temporale	43
3.2 Schemi di previsione	44
3.3 Iperparametri	47
3.4 Metriche predittive	49
3.5 Analisi dei parametri stimati	50
3.5.1 Effetto casa	50
3.5.2 <i>Ranking</i>	51
3.5.3 Evoluzione delle forze latenti	56
3.6 Analisi predittiva	62
3.6.1 Confronto con i modelli <i>goal-based</i>	62
3.6.2 Validazione a finestra crescente	69
3.7 Diagnostica di convergenza	73
Conclusioni	75
Appendice	77
A.1 Confronto tra schemi di campionamento	77
A.2 Diagnostiche di convergenza complete	77
A.3 Analisi di sensitività	83
A.4 Evoluzione dei <i>rankings</i>	86

A.5	Pesi della componente <i>slab</i>	91
A.6	Codice R	93

Bibliografia	101
---------------------	------------

Introduzione

La previsione degli esiti delle partite di calcio costituisce un problema statistico di interesse consolidato, in particolare per società sportive, staff tecnici, dirigenti e operatori del mercato delle scommesse. La letteratura statistica affronta questo problema attraverso due approcci differenti. Un primo approccio consiste nel modellare il numero di reti realizzate dalle squadre, sfruttando l'informazione contenuta nel punteggio finale. In alternativa, si può considerare l'esito qualitativo dell'incontro, distinguendo tra vittoria della squadra di casa, pareggio e vittoria della squadra ospite. In entrambi i casi, al fine di prevedere un risultato, occorre stimare la competitività delle squadre coinvolte. Questa quantità latente è tuttavia influenzata da eventi legati al contesto calcistico che si susseguono nel tempo, rendendo dunque interessante considerare una dinamica di evoluzione dei rendimenti di ogni compagine.

In questo lavoro si propone un modello *probit* ordinale dinamico, nel quale il parametro di forza di ogni squadra evolve secondo un *random walk* a tempi discreti, ispirato in parte alla sezione di discussione di Macrì-Demartino et al. (2026). In particolare, si intende estendere ad un contesto di modellazione di risultati qualitativi la metodologia proposta dagli autori, pensata per rendere adattiva la dinamica dei parametri stimati.

La capacità predittiva del modello viene valutata su dati reali provenienti dai massimi campionati europei, e viene confrontata con le specificazioni proposte in Macrì-Demartino et al. (2026). Vengono inoltre analizzati i risultati da un punto di vista interpretativo, mettendo a confronto le specificazioni proposte in termini di *ranking* e di evoluzione dei parametri in corrispondenza di eventi reali critici.

La tesi è organizzata come segue. Il Capitolo 1 presenta la letteratura sui principali modelli statistici per i risultati calcistici e ne discute le caratteristiche rilevanti ai fini del problema considerato. Il Capitolo 2 introduce il modello proposto e ne descrive la formulazione e la procedura di inferenza. Il Capitolo 3 presenta l'analisi empirica e discute i risultati ottenuti in termini di stima e previsione.

Capitolo 1

Modelli statistici per l'analisi dei risultati calcistici

La letteratura statistica sulla modellazione dei risultati calcistici si distingue in due approcci principali, in base alla variabile risposta considerata. L'approccio *result-based* modella invece l'esito qualitativo dell'incontro (vittoria della squadra di casa, pareggio o vittoria della squadra ospite), mentre l'approccio *goal-based* assume come oggetto della modellazione il numero di reti realizzate dalle due squadre, dal quale le probabilità dei tre esiti si ottengono per aggregazione. Il primo rinuncia al dettaglio del punteggio, e con esso a parte dell'informazione potenzialmente utile a valutare la forza delle squadre, ma consente formulazioni più parsimoniose e non richiede alcuna assunzione sulla dipendenza fra i gol (Goddard & Asimakopoulos, 2004; Egidi & Torelli, 2021). Il secondo utilizza un'informazione più ricca e restituisce, oltre alle probabilità di vittoria, pareggio e sconfitta, anche quelle dei singoli punteggi, ma richiede di specificare una struttura di dipendenza fra le reti delle due squadre e associa a ciascuna squadra una coppia di parametri, di attacco e di difesa.

1.1 L'approccio *goal-based*

Il modello proposto da Maher (1982) costituisce la pietra miliare della letteratura statistica sul calcio. L'ipotesi fondamentale è che il numero di gol segnati dalle due squadre in una stessa partita segua due distribuzioni di Poisson condizionatamente indipendenti.

Si indichi con K il numero di squadre di un campionato e con n il numero complessivo di partite disputate. Per la partita $i = 1, \dots, n$, disputata fra la squadra di casa h_i e

quella ospite a_i , con $h_i \neq a_i \in \{1, \dots, K\}$, si denotino con X_{h_i} e Y_{a_i} rispettivamente le reti segnate dalla squadra di casa e dalla squadra ospite. Il modello assume

$$X_{h_i} \mid \lambda_{1,i} \sim \text{Poisson}(\lambda_{1,i}), \quad Y_{a_i} \mid \lambda_{2,i} \sim \text{Poisson}(\lambda_{2,i}), \quad X_{h_i} \perp Y_{a_i} \mid \lambda_{1,i}, \lambda_{2,i},$$

dove $\lambda_{1,i}$ e $\lambda_{2,i}$ rappresentano i tassi attesi di reti segnate, rispettivamente, dalla squadra di casa e dalla squadra ospite nella partita i . Tali tassi sono modellati mediante la seguente struttura log-lineare

$$\log \lambda_{1,i} = \beta_0 + \alpha + \beta_{h_i}^{(\text{att})} + \beta_{a_i}^{(\text{dif})}, \quad \log \lambda_{2,i} = \beta_0 + \beta_{a_i}^{(\text{att})} + \beta_{h_i}^{(\text{dif})},$$

dove β_0 rappresenta un'intercetta comune a tutte le partite, α l'effetto sistematico di giocare in casa, e $\beta_k^{(\text{att})}$ e $\beta_k^{(\text{dif})}$ rappresentano rispettivamente la capacità offensiva e difensiva della squadra k .

Dixon & Coles (1997) apportano una modifica al modello base di Maher (1982) circa l'assunzione di indipendenza condizionata fra i gol delle due squadre. Confrontando le frequenze osservate dei risultati con quelle previste dal modello, gli autori riscontrano una sottostima dei pareggi per 0-0 e 1-1 e una sovrastima delle vittorie di misura. Per correggere questa discrepanza, gli autori introducono un fattore $\tau(x_i, y_i; \rho)$, che perturba la probabilità congiunta prevista dal modello originale, limitatamente alle quattro combinazioni di punteggio più basse:

$$P(X_{h_i} = x_i, Y_{a_i} = y_i) = \tau(x_i, y_i; \rho) \cdot \frac{e^{-\lambda_{1,i}} \lambda_{1,i}^{x_i}}{x_i!} \cdot \frac{e^{-\lambda_{2,i}} \lambda_{2,i}^{y_i}}{y_i!},$$

con

$$\tau(x_i, y_i; \rho) = \begin{cases} 1 - \rho \lambda_{1,i} \lambda_{2,i} & \text{se } x_i = 0, y_i = 0, \\ 1 + \rho \lambda_{1,i} & \text{se } x_i = 0, y_i = 1, \\ 1 + \rho \lambda_{2,i} & \text{se } x_i = 1, y_i = 0, \\ 1 - \rho & \text{se } x_i = 1, y_i = 1, \\ 1 & \text{altrimenti.} \end{cases}$$

Il parametro ρ governa quindi segno e intensità della dipendenza fra i punteggi bassi, e per $\rho = 0$ il modello si riduce alla specificazione base di Maher (1982).

Un approccio alternativo che considera la dipendenza tra i punteggi è proposto da Karlis & Ntzoufras (2003), che generalizzano il modello di Maher (1982) introducendo un parametro di covarianza tra X_{h_i} e Y_{a_i} .

L'assunzione alla base della distribuzione di Poisson è l'uguaglianza tra media e

varianza, ipotesi che non regge necessariamente nei dati calcistici reali. Reep et al. (1971) osservano come la distribuzione dei punteggi nel calcio si discosti da quella teorica della Poisson, evidenziando la sovradisersione. Una soluzione possibile è quella di adottare una distribuzione binomiale negativa per il conteggio delle reti, che mantiene la stessa struttura log-lineare del modello di Maher per il valore atteso dei gol e introduce un parametro aggiuntivo di sovradisersione:

$$X_{h_i} \mid \lambda_{1,i}, \phi_1 \sim \text{NB}(\lambda_{1,i}, \phi_1), \quad Y_{a_i} \mid \lambda_{2,i}, \phi_2 \sim \text{NB}(\lambda_{2,i}, \phi_2),$$

con densità discreta

$$P(X_{h_i} = x_i) = \frac{\Gamma(x_i + \phi_1)}{\Gamma(x_i + 1) \Gamma(\phi_1)} \left(\frac{\phi_1}{\lambda_{1,i} + \phi_1} \right)^{\phi_1} \left(\frac{\lambda_{1,i}}{\lambda_{1,i} + \phi_1} \right)^{x_i},$$

dove $\mathbb{E}(X_{h_i}) = \lambda_{1,i}$ e $\text{Var}(X_{h_i}) = \lambda_{1,i} + \lambda_{1,i}^2/\phi_1$. In modo analogo si definisce anche la densità marginale di Y_{a_i} .

Un approccio diverso, proposto da Karlis & Ntzoufras (2009), consiste nel modellare la differenza di reti segnate dalle due squadre mediante la distribuzione di Skellam (Skellam, 1946), definita come differenza fra due variabili casuali di Poisson indipendenti $Z_i = X_{h_i} - Y_{a_i}$, con funzione di probabilità

$$P(Z_i = z) = \exp\{-(\lambda_{1,i} + \lambda_{2,i})\} \left(\frac{\lambda_{1,i}}{\lambda_{2,i}} \right)^{z/2} I_z\left(2\sqrt{\lambda_{1,i}\lambda_{2,i}}\right), \quad z \in \mathbb{Z},$$

dove $I_z(\cdot)$ è la funzione di Bessel modificata di ordine z , $\mathbb{E}(Z_i) = \lambda_{1,i} - \lambda_{2,i}$ e $\text{Var}(Z_i) = \lambda_{1,i} + \lambda_{2,i}$. Come osservano Karlis & Ntzoufras (2009), questa distribuzione della differenza reti cattura così sia la sovradisersione sia la dipendenza intrinseca fra i punteggi delle due squadre, senza richiederne una modellazione esplicita.

Sia il modello di Poisson bivariato sia quello di Skellam ammettono una variante che corregge la tendenza dei modelli a sottostimare la frequenza dei pareggi. Karlis & Ntzoufras (2005) propongono una versione a inflazione diagonale del modello Poisson bivariato, che aggiunge massa di probabilità a tutte le celle corrispondenti ad un punteggio di pareggio; Karlis & Ntzoufras (2009) invece adottano una versione a inflazione di zeri della distribuzione di Skellam, che aggiunge massa di probabilità unicamente in corrispondenza del valore nullo della differenza reti.

Un'ulteriore variante del modello di Maher è proposta da Baio & Blangiardo (2010), che formulano un modello gerarchico bayesiano, adottando distribuzioni a priori normali per le abilità offensive e difensive.

Tutte le metodologie descritte finora assumono che i parametri relativi alle abilità delle squadre rimangano costanti nel tempo, un'ipotesi che risulta opinabile in questo contesto. Alcuni fenomeni, quali infortuni, cambi di allenatore, trasferimenti e cambiamenti nella struttura delle rose rendono le prestazioni delle squadre una quantità che varia nel tempo. Una prima estensione dinamica formale dei modelli *goal-based* è proposta da Rue & Salvesen (2000), che modellano l'evoluzione delle abilità in tempo continuo tramite un processo gaussiano la cui varianza cresce con il tempo trascorso fra le partite. Koopman & Lit (2019) considerano le abilità delle squadre come vettori di stato che evolvono nel tempo, adattando il modello di Poisson bivariato ad un approccio *state-space*. Owen (2011) propone un'alternativa a tempi discreti in cui le abilità offensive e difensive di ciascuna squadra evolvono di giornata in giornata secondo un processo *random walk* gaussiano,

$$\beta_{k,t}^{(\text{att})} \mid \beta_{k,t-1}^{(\text{att})} \sim \mathcal{N}\left(\beta_{k,t-1}^{(\text{att})}, \sigma^2\right), \quad \beta_{k,t}^{(\text{dif})} \mid \beta_{k,t-1}^{(\text{dif})} \sim \mathcal{N}\left(\beta_{k,t-1}^{(\text{dif})}, \sigma^2\right),$$

dove t indicizza le giornate del campionato. Al fine di ridurre la complessità del modello, la varianza di evoluzione è unica, comune a tutte le squadre e condivisa fra le abilità di attacco e difesa, penalizzando dunque la flessibilità del modello quando le capacità offensive e difensive di una stessa squadra evolvono a ritmi diversi, o quando alcuni periodi temporali richiederebbero un aggiornamento più marcato di altri.

Egidi et al. (2018) adottano una versione a tempi discreti dell'approccio di Rue & Salvesen (2000), arricchita dalle quote dei *bookmakers* come ulteriore fonte informativa. A differenza di Owen (2011), le varianze di evoluzione sono distinte per le abilità offensive e difensive, sebbene ancora comuni a tutte le squadre.

Macrì-Demartino et al. (2026) propongono infine una generalizzazione dell'approccio di Owen (2011) tramite distribuzioni a priori sui parametri specifiche per squadra, tipo di abilità e periodo temporale:

$$\beta_{k,t}^{(\text{att})} \mid \beta_{k,t-1}^{(\text{att})}, \varphi_{k,t}^{(\text{att})} \sim \mathcal{N}\left(\beta_{k,t-1}^{(\text{att})}, 1/(\varphi_{k,t}^{(\text{att})})^2\right),$$

e analogamente per $\beta_{k,t}^{(\text{dif})}$. La distribuzione a priori della precisione di evoluzione $\varphi_{k,t}^{(\text{att})}$ è specificata mediante una distribuzione *spike-and-slab*, coerentemente con l'approccio di Hobbs et al. (2012) descritto nel dettaglio nel paragrafo 2.4. Questa specificazione permette al modello di reagire rapidamente ai cambiamenti nelle prestazioni di una squadra e di restare stabile quando le prestazioni sono regolari, con miglioramenti sistematici delle capacità predittive rispetto ai modelli dinamici tradizionali.

1.2 L'approccio *result-based*

Un approccio alternativo ai modelli *goal-based* consiste nel modellare direttamente l'esito della partita come variabile risposta qualitativa. I modelli che adottano questa impostazione, indicati come modelli *result-based*, possono essere distinti in funzione del trattamento dell'esito: esso può essere considerato una variabile categoriale nominale, senza assumere alcuna relazione d'ordine tra le modalità, oppure una variabile ordinale, introducendo una struttura di ordinamento tra gli esiti.

Tra i modelli nominali rientra il modello di Bradley–Terry (Bradley & Terry, 1952), originariamente sviluppato per i confronti a coppie. A ciascun concorrente k viene associato un parametro di forza $\psi_k \in \mathbb{R}$ e, nel caso di due soli esiti possibili, la probabilità che il concorrente k prevalga sul concorrente k' è

$$P(k \text{ vince contro } k') = \frac{\exp(\psi_k)}{\exp(\psi_k) + \exp(\psi_{k'})}.$$

Davidson (1970) propone un'estensione del modello Bradley–Terry che permette di contemplare anche l'esito di pareggio. Un'ulteriore applicazione di un modello a risposta nominale è proposta da Nyberg (2014), che utilizza una regressione logistica multinomiale per analizzare l'efficienza del mercato delle scommesse, impiegando come covariate le probabilità implicite medie desunte dalle quote dei *bookmakers*. Egidi & Torelli (2021) adottano una regressione multinomiale bayesiana, nella quale i predittori lineari per la vittoria in casa e in trasferta dipendono da parametri di attacco e difesa per squadra, analoghi a quelli dei modelli *goal-based*.

Un'impostazione differente consiste nel riconoscere una struttura ordinale nell'esito della partita. In questa prospettiva, Kuypers (2000) sottolinea l'opportunità di ricorrere a un modello di regressione ordinale, poiché vittoria, pareggio e sconfitta possono essere collocati lungo una scala associata alla forza relativa delle due squadre. Un contributo in questa direzione è quello di Koning (2000), che modella direttamente i risultati del calcio professionistico olandese mediante un modello *probit* ordinale (Agresti, 2002, Capitolo 7). L'esito della partita è determinato dal vantaggio di giocare in casa e dalla differenza di qualità tra le due squadre, rappresentata da un unico parametro di forza per ciascuna squadra. Un'impostazione affine è adottata da Goddard & Asimakopoulou (2004), che utilizzano un modello *probit* ordinale per prevedere gli esiti delle partite del campionato inglese, arricchito da diverse covariate come ad esempio l'importanza della partita ai fini della classifica finale, allo scopo di analizzare l'efficienza del mercato delle scommesse a quota fissa.

Per quanto riguarda le estensioni dinamiche dei modelli *result-based*, Knorr-Held

(2000) propone un modello a funzione di legame cumulativa per gli esiti ordinati, nel quale i parametri di abilità delle squadre evolvono nel tempo secondo un processo *random walk*. La dinamica delle abilità è trattata come un problema di lisciamiento, con inferenza condotta mediante filtro di Kalman. Cattelan et al. (2013) propongono un'estensione dinamica del modello di Bradley–Terry in cui le forze di casa e trasferta di ciascuna squadra evolvono secondo processi a media mobile a pesi esponenziali; i pareggi sono incorporati tramite una funzione di legame cumulativa (Agresti, 2002, Capitolo 7), dove l'esito qualitativo deriva dal confronto di una variabile latente con due soglie vincolate ad essere simmetriche. Il modello, stimato mediante massima verosimiglianza e applicato alla Serie A italiana, mostra un miglioramento della capacità predittiva rispetto alla versione statica.

1.3 Motivazione e proposta metodologica

La rassegna nei paragrafi precedenti mostra che le estensioni dinamiche, con le quali si è cercato di superare l'ipotesi di forza costante, si sono sviluppate in larga parte nell'ambito *goal-based*, dove a ciascuna squadra sono associati due parametri distinti di attacco e difesa. Nel filone *result-based*, invece, la modellazione dinamica è stata esplorata meno sistematicamente, benché in quel contesto la forza di una squadra sia rappresentata da un unico parametro. Un obiettivo di questo lavoro è quello di modellare efficacemente i risultati calcistici attraverso un modello ordinale dinamico fondato su un solo parametro di forza per squadra, ottenendo uno strumento parsimonioso, interpretabile e flessibile, capace di produrre automaticamente una graduatoria delle squadre e adeguato alla previsione degli esiti qualitativi. Un unico indice per squadra è direttamente leggibile come livello competitivo complessivo, pertanto se ne possono seguire le traiettorie nel tempo, ricavare in ciascun periodo un ordinamento e misurare la diversa variabilità dei rendimenti, in linea con l'impiego di tali parametri nei modelli per confronti a coppie (Cattelan, 2012) e con l'esempio applicato al calcio di Macrì Demartino et al. (2025). Con parametri distinti di attacco e difesa, al contrario, la costruzione di una graduatoria richiede un criterio ulteriore per combinarli.

Si considera a tale scopo un modello *probit* ordinale in cui la forza di ciascuna squadra è rappresentata da un unico parametro, come in Koning (2000), reso dinamico mediante un processo *random walk* gaussiano, analogamente ai modelli di Knorr-Held (2000), Rue & Salvesen (2000) e Owen (2011). Una limitazione degli approcci dinamici citati è la specificazione di una varianza di innovazione fissata, che risulta poco accomodante rispetto ai fenomeni che possono caratterizzare il contesto in esame. Alcune

squadre possono presentare rendimenti molto stabili nel tempo, ma in corrispondenza di eventi critici subire dei cambiamenti anche radicali nelle loro prestazioni. La proposta dinamica adattiva di Macrì-Demartino et al. (2026) si pone l'obiettivo di considerare proprio questo tipo di eventi nella specificazione del modello. La struttura presentata consente in particolare di modulare la dinamicità del modello, favorendo una maggiore stabilità temporale oppure una maggiore reattività ai dati recenti. In questo caso, il modello bayesiano viene stimato tramite metodi *Hamiltonian Monte Carlo*. Un'attenta formulazione delle distribuzioni a priori e della struttura gerarchica potrebbe tuttavia permettere di recuperare distribuzioni *full conditional* in forma chiusa consentendo l'implementazione di un algoritmo di *Gibbs sampling*, alleggerendo così notevolmente il costo computazionale.

Il contributo del lavoro si articola su due piani. Sul piano applicato, si intende costruire un modello dinamico dotato di buone capacità previsive sui risultati calcistici e in grado di restituire, come sottoprodotto della stima, una graduatoria delle squadre e una lettura della loro evoluzione temporale. Sul piano metodologico, l'interesse riguarda la modellazione flessibile della dinamica dei parametri in un modello per dati ordinali e la corrispondente proposta di inferenza.

Capitolo 2

Modellazione

In questo capitolo viene presentato un modello bayesiano dinamico a risposta ordinale per l'analisi dei risultati calcistici. L'obiettivo è stimare un parametro di forza latente per ciascuna squadra, che ne descriva la capacità relativa rispetto alle altre, e utilizzare tali stime per prevedere l'esito di una partita, trattato come una variabile categoriale ordinata. La natura dinamica del modello permette alla forza delle squadre di evolvere nel tempo, così da cogliere le variazioni di rendimento che possono manifestarsi nel corso delle stagioni. Di questa struttura si considerano due specificazioni alternative, distinte dal modo in cui è modellata la variabilità dell'evoluzione temporale delle forze, affiancate da una versione statica impiegata come termine di riferimento.

Nel paragrafo 2.1 viene specificato il modello statistico, nel paragrafo 2.2 vengono discusse le distribuzioni a priori dei parametri e introdotte le specificazioni poste a confronto, nel paragrafo 2.3 si richiamano i fondamenti dell'inferenza bayesiana e dei metodi *Markov chain Monte Carlo*, ed infine nel paragrafo 2.4 vengono derivate le distribuzioni *full conditional* necessarie per implementare l'algoritmo *Gibbs sampling* di stima.

2.1 Specificazione del modello

Per descrivere il modello, si introduce la notazione utilizzata nel seguito. Si denotano con n il numero totale di partite osservate, con K il numero di squadre partecipanti e con T il numero di periodi temporali in cui la stagione è suddivisa. L'indice $i = 1, \dots, n$ individua una partita, l'indice $k = 1, \dots, K$ una squadra e l'indice $t = 1, \dots, T$ un periodo temporale.

Per ogni partita i , la coppia ordinata (h_i, a_i) individua rispettivamente la squadra di casa e la squadra ospite, con

$$(h_i, a_i) \in \mathcal{S} = \{(k, k') : k, k' \in \{1, \dots, K\}, k \neq k'\}.$$

Inoltre, sia $t_i \in \{1, \dots, T\}$ il periodo temporale di appartenenza della partita. Per ciascuna squadra k e periodo t , si definisce l'insieme

$$\mathcal{I}_t^{(k)} = \{i : t_i = t, h_i = k\} \cup \{i : t_i = t, a_i = k\},$$

che raccoglie tutte le partite disputate dalla squadra k nel periodo t . La sua cardinalità è denotata con $n_{k,t} = |\mathcal{I}_t^{(k)}|$. Analogamente, si definisce

$$\mathcal{I}_t = \{i : t_i = t\},$$

l'insieme di tutte le partite disputate nel periodo t , mentre il numero complessivo di partite appartenenti a tale periodo è indicato da $n_t = |\mathcal{I}_t|$. Per ogni partita $i = 1, \dots, n$ si osserva il risultato $Y_i \in \{-1, 0, 1\}$, dove -1 indica la vittoria della squadra ospite, 0 il pareggio e 1 la vittoria della squadra di casa.

La natura ordinale della variabile risposta suggerisce l'impiego di un modello per risposte categoriali ordinate. In particolare, si adotta una funzione di legame *probit* (Agresti, 2002, Capitolo 7). Tale formulazione risulta particolarmente conveniente in ambito bayesiano, in quanto consente di sfruttare la tecnica di *data augmentation*, proposta da Tanner & Wong (1987) e successivamente applicata nel contesto della modellazione di risposte categoriali da Albert & Chib (1993). Questo approccio, con un'opportuna scelta delle distribuzioni a priori, consente di ottenere distribuzioni condizionate complete (*full conditional distributions*) in forma chiusa, rendendo possibile l'impiego di un algoritmo di *Gibbs sampling* (Geman & Geman, 1984; Gelfand & Smith, 1990), descritto ad esempio in Gelman et al. (2014, paragrafo 11.1).

L'idea proposta da Albert & Chib (1993) consiste nel riformulare il modello *probit* introducendo una variabile latente continua per ciascuna osservazione, trattata come un dato non osservato (*data augmentation*). In questa formulazione, la risposta categoriale osservata è determinata esclusivamente dall'intervallo, individuato da opportune soglie, entro cui ricade la variabile latente. Se quest'ultima fosse osservabile, il problema si ridurrebbe a un ordinario modello di regressione lineare normale, per il quale la distribuzione a posteriori dei parametri è facilmente ottenibile. Poiché la variabile latente non è osservata, Albert & Chib (1993) mostrano che la sua distribuzione condizionata alla risposta osservata e ai parametri del modello è una distribuzione normale troncata.

Questo risultato consente di costruire un algoritmo di *Gibbs sampling* che alterna il campionamento delle variabili latenti e dei parametri del modello, evitando il calcolo diretto della distribuzione a posteriori, la quale non risulta disponibile in forma chiusa.

Nel contesto specifico di questo lavoro si assume pertanto che il risultato della partita sia generato da una variabile latente gaussiana, la cui media dipende dalla differenza di forza tra le due squadre coinvolte nell'incontro. La risposta osservata è ottenuta confrontando tale variabile con un insieme di soglie fissate, γ . Per ogni partita $i = 1, \dots, n$ si introduce una variabile latente continua $Z_i \in \mathbb{R}$ tale che:

$$Z_i = \eta_i + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2), \quad (2.1)$$

con

$$\eta_i = \alpha_{t_i} + \beta_{h_i, t_i} - \beta_{a_i, t_i}, \quad (2.2)$$

dove $\alpha_t \in \mathbb{R}$ rappresenta l'effetto casa del periodo t , comune a tutte le partite di quel periodo e indipendente dalle squadre coinvolte; $\beta_{k,t} \in \mathbb{R}$ è la forza della squadra k nel periodo t , che compare con segno positivo quando la squadra gioca in casa e con segno negativo quando gioca in trasferta; $\sigma_\varepsilon^2 \in \mathbb{R}_+$ è la varianza del termine d'errore ε_i .

Per quanto riguarda l'identificabilità del modello, si osservi che le probabilità degli esiti dipendono dai parametri soltanto attraverso il rapporto $(\gamma_j - \eta_i)/\sigma_\varepsilon$. Tale rapporto è invariante rispetto a traslazioni di η_i e delle soglie per una medesima costante, nonché a riscalamanti di η_i , delle soglie e di σ_ε per un fattore positivo. Sono pertanto necessarie due restrizioni, l'una di posizione e l'altra di scala, che si sceglie qui di imporre fissando entrambe le soglie,

$$\gamma_1 = -0.5, \quad \gamma_2 = 0.5,$$

e lasciando libera la varianza dell'errore σ_ε^2 . La parametrizzazione comunemente adottata in letteratura per il modello *probit* ordinale procede in senso opposto, fissando la varianza dell'errore e una delle soglie, trattando le rimanenti come parametri da stimare, come in Albert & Chib (1993) e, in ambito calcistico, in Koning (2000). I due approcci impongono lo stesso numero di vincoli e individuano il medesimo modello statistico, conducendo, a meno di una trasformazione affine, alle stesse probabilità stimate. La scelta adottata è motivata dall'interpretazione della variabile latente Z_i come approssimazione continua della differenza di reti della partita. In tale prospettiva, γ_1 e γ_2 si collocano nei punti medi fra i valori -1 e 0 e fra i valori 0 e 1 della differenza di reti, separando dunque le regioni associate a una differenza negativa, nulla e positiva. La simmetria delle soglie rispetto all'origine, unitamente alla simmetria della distribuzione normale dell'errore,

garantisce che, a parità di forza fra le due squadre e in assenza di effetto casa, le probabilità di vittoria casalinga e di vittoria in trasferta coincidano, cosicché ogni asimmetria fra i due esiti nel modello stimato risulti attribuibile esclusivamente ai parametri α e β . La parametrizzazione adottata presenta infine un vantaggio dal punto di vista computazionale. Cowles (1996) mostra che, nei modelli *probit* ordinali, le variabili latenti e le soglie possono risultare fortemente correlate a posteriori, rendendo poco efficiente il loro aggiornamento separato all'interno di un algoritmo di *Gibbs sampling*, e propone pertanto un aggiornamento congiunto mediante un passo Metropolis–Hastings. Poiché nel presente modello le soglie sono fissate, tale fonte di dipendenza non si presenta, e l'algoritmo può fondarsi esclusivamente su aggiornamenti in forma chiusa.

Definendo $-\infty = \gamma_0 < \gamma_1 = -0.5 < \gamma_2 = 0.5 < \gamma_3 = +\infty$, il legame tra la variabile latente gaussiana e la risposta osservata è

$$Y_i = j \iff \gamma_{j+1} < Z_i \leq \gamma_{j+2}, \quad j \in \{-1, 0, 1\}. \quad (2.3)$$

La probabilità di ciascun esito è espressa in termini della funzione di ripartizione della normale standard $\Phi(\cdot)$:

$$P(Y_i = j \mid \eta_i, \sigma_\varepsilon^2) = \Phi\left(\frac{\gamma_{j+2} - \eta_i}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\gamma_{j+1} - \eta_i}{\sigma_\varepsilon}\right), \quad j \in \{-1, 0, 1\},$$

ossia

$$\begin{aligned} P(Y_i = -1 \mid \eta_i, \sigma_\varepsilon^2) &= \Phi\left(\frac{\gamma_1 - \eta_i}{\sigma_\varepsilon}\right), \\ P(Y_i = 0 \mid \eta_i, \sigma_\varepsilon^2) &= \Phi\left(\frac{\gamma_2 - \eta_i}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\gamma_1 - \eta_i}{\sigma_\varepsilon}\right), \\ P(Y_i = 1 \mid \eta_i, \sigma_\varepsilon^2) &= 1 - \Phi\left(\frac{\gamma_2 - \eta_i}{\sigma_\varepsilon}\right). \end{aligned}$$

Si noti inoltre che, poiché il predittore lineare (2.2) dipende dalle forze latenti solo attraverso la differenza $\beta_{h_i, t_i} - \beta_{a_i, t_i}$, queste sono identificate soltanto a meno di una costante additiva comune in ciascun periodo. Per garantirne l'identificabilità si impone il vincolo a somma zero $\sum_{k=1}^K \beta_{k, t} = 0$ per ogni t .

La Figura 2.1 illustra il meccanismo a variabile latente descritto in questo paragrafo.

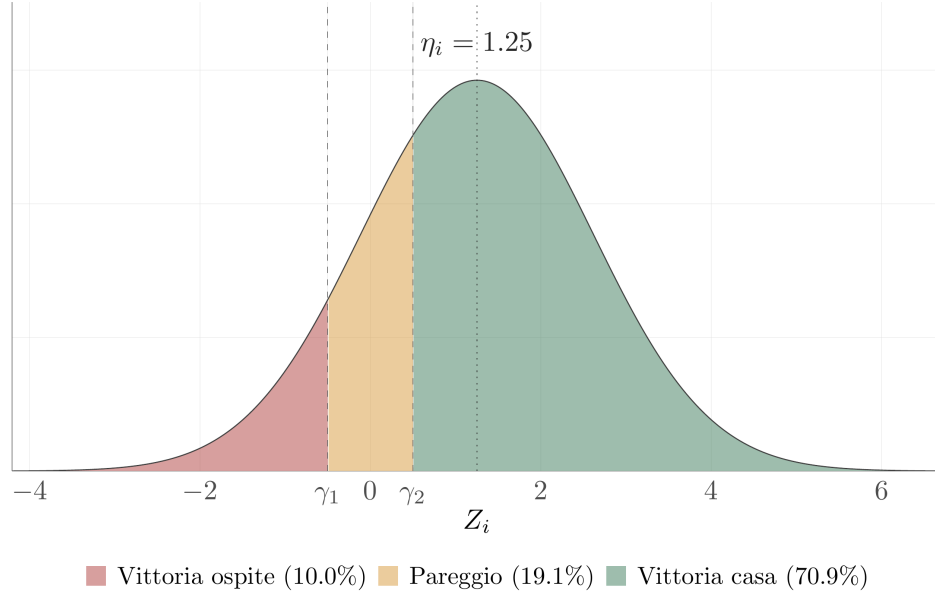


FIGURA 2.1: Esempio illustrativo del meccanismo a variabile latente del modello, applicato alla partita Milan-Sassuolo (1-0) del 30/12/2023, valida per la diciottesima giornata di Serie A. La curva rappresenta la densità di $Z_i \sim N(\eta_i, \sigma_\varepsilon^2)$, con η_i e σ_ε^2 fissati alle rispettive medie a posteriori stimate dal modello con varianza di evoluzione *spike-and-slab* squadra-specifica (paragrafo 2.4); le aree colorate corrispondono alle probabilità dei tre esiti, delimitate dalle soglie fisse $\gamma_1 = -0.5$ e $\gamma_2 = 0.5$.

2.2 Distribuzioni a priori dei parametri

La specificazione del modello bayesiano richiede l'assegnazione di una distribuzione a priori a ciascun parametro introdotto nel paragrafo 2.1. Le scelte effettuate sono volte a riflettere l'evoluzione temporale delle forze delle squadre attraverso una struttura dinamica di tipo *random walk*, in conformità con l'approccio adottato in letteratura (Rue & Salvesen, 2000; Owen, 2011; Macrì-Demartino et al., 2026). I parametri del modello includono le forze latenti $\beta = (\beta_{1,1}, \dots, \beta_{K,T}) \in \mathbb{R}^{K \times T}$, le varianze di evoluzione $\sigma^2 = (\sigma_2^2, \dots, \sigma_T^2) \in \mathbb{R}_+^{T-1}$, la varianza dell'errore $\sigma_\varepsilon^2 \in \mathbb{R}_+$ e l'effetto casa $\alpha = (\alpha_1, \dots, \alpha_T) \in \mathbb{R}^T$.

Per le forze latenti delle squadre si adottano distribuzioni a priori dinamiche attraverso un *random walk* gaussiano. Per il periodo iniziale

$$\beta_{k,1} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_1^2), \quad k = 1, \dots, K,$$

con $\sigma_1^2 > 0$ fissato affinché la priori sia debolmente informativa, la calibrazione adottata

e la sua motivazione sono discusse nel paragrafo 3.3. Per i periodi successivi si assume

$$\beta_{k,t} \mid \beta_{k,t-1}, \sigma_t^2 \sim \mathcal{N}(\beta_{k,t-1}, \sigma_t^2), \quad k = 1, \dots, K, \quad t = 2, \dots, T.$$

La forza di ogni squadra al periodo t ha una distribuzione a priori centrata sul valore del periodo precedente. Il parametro σ_t^2 governa quanto la forza di una squadra può cambiare da un periodo all'altro: valori grandi permettono maggiore incertezza, valori piccoli implicano un rendimento più stabile e uno *shrinkage* più forte verso il valore del periodo precedente.

In questo lavoro vengono considerate due proposte di specificazione del modello, che differiscono nella distribuzione a priori e nella dimensionalità di σ_t^2 , affiancate da un modello statico impiegato come termine di riferimento. In quest'ultimo, indicato nel seguito con SOP (*Static Ordered Probit*), la varianza di innovazione è nulla e la forza di ciascuna squadra resta costante nel tempo, $\beta_{k,t} = \beta_{k,1}$ per ogni t .

La prima proposta, il modello dinamico DOP (*Dynamic Ordered Probit*), assume una singola varianza di innovazione comune a tutte le squadre in ciascun periodo,

$$\sigma_t^2 \stackrel{\text{iid}}{\sim} \text{IGa}(a, b), \quad t = 2, \dots, T, \quad (2.4)$$

con iperparametri $a, b > 0$ fissati. La distribuzione gamma inversa è particolarmente indicata in questo contesto in quanto mantiene la coniugatezza con il modello normale (Gelman et al., 2014, paragrafo 2.6).

La seconda proposta, il modello dinamico adattivo DAOP (*Dynamic Adaptive Ordered Probit*), ispirata a Macrì-Demartino et al. (2026), introduce una varianza di innovazione specifica per squadra e periodo. Questa specificazione permette di rilevare comportamenti eterogenei nell'evoluzione temporale delle prestazioni delle squadre, aumentando tuttavia in modo considerevole il numero di parametri da stimare. Eventi come un cambio di allenatore o una sessione di calciomercato possono causare variazioni repentine nelle prestazioni di alcune squadre, mentre altre possono mantenere un livello pressoché invariato nel tempo. Una varianza comune impone lo stesso grado di adattamento a tutte le squadre all'interno di un medesimo periodo. Ciò può determinare un eccessivo trasferimento di informazione storica (*over-borrowing*) per le squadre soggette a cambiamenti significativi oppure, al contrario, un trasferimento insufficiente (*under-borrowing*) per quelle la cui evoluzione rimane stabile. Per affrontare questo problema, Macrì-Demartino et al. (2026) adottano un approccio basato sulle *commensurate priors* introdotte da Hobbs et al. (2011, 2012). Queste distribuzioni regolano il trasferimento dell'informazione storica in funzione della sua coerenza (*commensurability*) con i

dati correnti. Il parametro di interesse θ viene centrato su una stima storica θ_0 tramite $\theta \mid \theta_0, \psi^2 \sim \mathcal{N}(\theta_0, 1/\psi^2)$, dove ψ^2 è il parametro di commensurabilità. Quando i dati storici e quelli correnti risultano coerenti, la priori contrae θ verso θ_0 , migliorando l'efficienza della stima; quando invece emerge un disaccordo sostanziale, la priori esercita un'influenza minima su θ (Murray et al., 2015).

Hobbs et al. (2012) propongono una distribuzione a priori *spike-and-slab* (Mitchell & Beauchamp, 1988) per ψ^2 , che consiste in una mistura a due componenti. In questa struttura, una componente (*spike*) concentra la massa di probabilità su valori elevati di ψ^2 , promuovendo una forte contrazione di θ verso θ_0 , mentre l'altra componente (*slab*) è distribuita su valori più bassi, permettendo a θ di discostarsi da θ_0 . Macrì-Demartino et al. (2026) adottano una versione continua di questa mistura, proposta da Hong et al. (2018), che si basa su distribuzioni normali troncate. Nel presente studio, si applica la stessa idea alla definizione della distribuzione a priori della varianza $\sigma_{k,t}^2$, anziché alla precisione, utilizzando una mistura di due distribuzioni gamma inverse, invertendo coerentemente il ruolo dello *spike* e della *slab*. In dettaglio, per $k = 1, \dots, K$, $t = 2, \dots, T$, si assume

$$\sigma_{k,t}^2 \sim p_k \text{IGa}(a_{\text{sp}}, b_{\text{sp}}) + (1 - p_k) \text{IGa}(a_{\text{sl}}, b_{\text{sl}}), \quad (2.5)$$

dove gli iperparametri $a_{\text{sp}}, b_{\text{sp}}, a_{\text{sl}}, b_{\text{sl}} > 0$ caratterizzano rispettivamente lo *spike* e la *slab*, mentre $p_k \in [0, 1]$ è il peso della mistura, specifico per ciascuna squadra k . L'interpretazione e la stima di p_k sono discusse nel paragrafo 2.4. Lo *spike* concentra la massa di probabilità su valori piccoli di $\sigma_{k,t}^2$, favorendo una forte persistenza temporale del coefficiente; la *slab*, invece, è concentrata su valori più elevati della varianza e permette variazioni più marcate tra periodi consecutivi, limitando il trasferimento di informazione storica. La Figura 2.2 illustra la distribuzione e il meccanismo di *shrinkage* adattivo sui parametri di forza delle squadre.

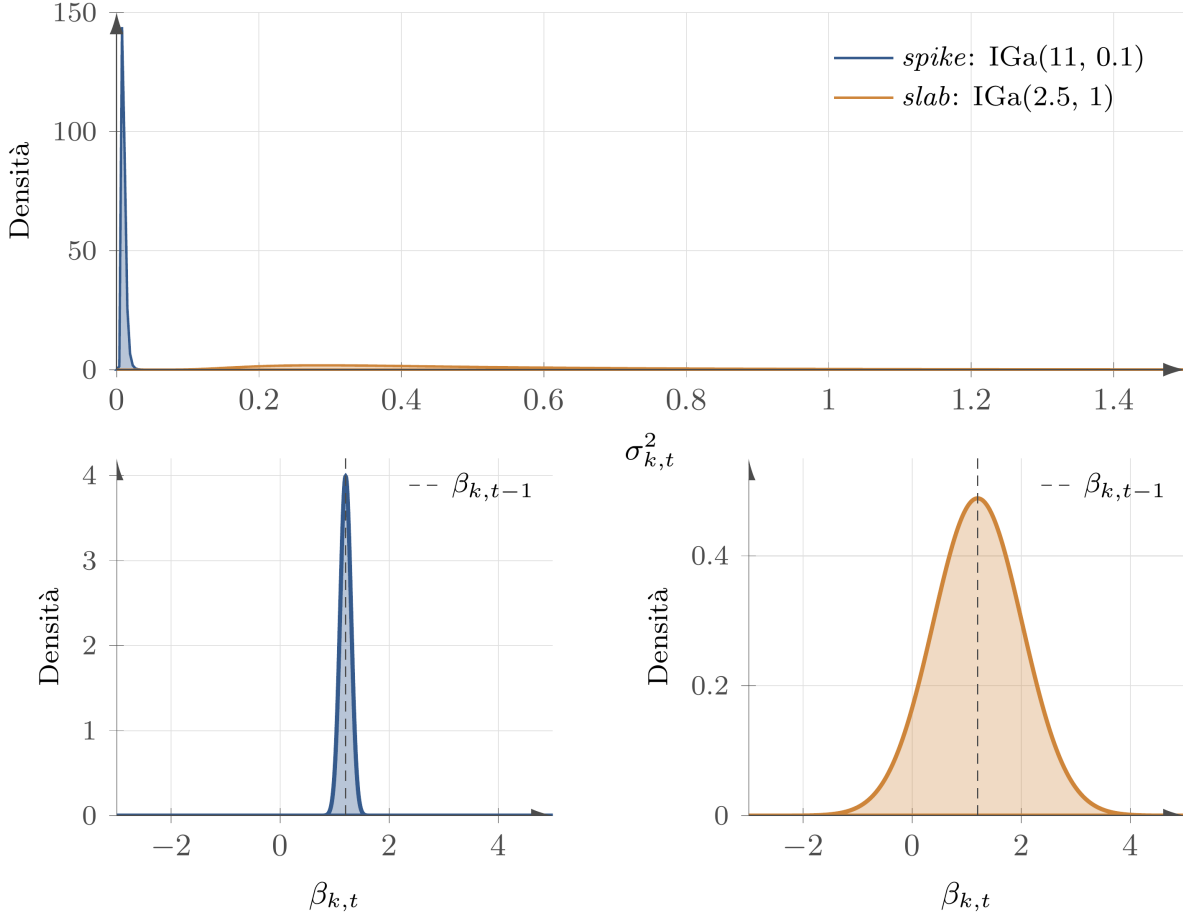


FIGURA 2.2: Meccanismo di commensurabilità adattivo indotto dalla distribuzione di $\sigma^2_{k,t}$. Pannello superiore: le due componenti della mistura. Pannello inferiore: distribuzione a priori per $\beta_{k,t}$ indotta dalle due componenti, centrata sul valore del periodo precedente $\beta_{k,t-1}$ (linea tratteggiata).

Per la varianza dell'errore della variabile latente gaussiana si adotta, analogamente alla (2.4), una priori gamma inversa per sfruttare la coniugatezza con il modello normale

$$\sigma_\varepsilon^2 \sim \text{IGa}(a_\varepsilon, b_\varepsilon),$$

con iperparametri $a_\varepsilon, b_\varepsilon > 0$ fissati.

Il parametro dell'effetto casa è specifico per periodo, senza imporre alcun legame fra periodi successivi

$$\alpha_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2), \quad t = 1, \dots, T,$$

con $\sigma_\alpha^2 > 0$ fissato affinché la priori sia debolmente informativa. Trattare l'effetto casa come indipendente fra periodi, anziché vincolarlo a un unico valore comune, permette al modello di adattarsi a variazioni sistematiche del vantaggio del giocare in casa fra un

periodo e l'altro, come il crollo osservato nella stagione 2020/21 giocata a porte chiuse (paragrafo 3.5.1).

2.3 Inferenza bayesiana tramite metodi Monte Carlo

Nell'approccio bayesiano, l'inferenza sui parametri del modello si basa sulla distribuzione a posteriori, ottenuta combinando l'informazione contenuta nei dati con quella espressa dalle distribuzioni a priori mediante il teorema di Bayes. Indicando con θ il vettore dei parametri, appartenente allo spazio parametrico Θ , e con y i dati osservati, la distribuzione a posteriori è definita come

$$\pi(\theta | y) = \frac{p(y | \theta) \pi(\theta)}{\int_{\Theta} p(y | \theta) \pi(\theta) d\theta} \propto p(y | \theta) \pi(\theta), \quad (2.6)$$

dove $p(y | \theta)$ rappresenta la verosimiglianza e $\pi(\theta)$ la distribuzione a priori. Il denominatore della (2.6) è la costante di normalizzazione, necessaria affinché $\pi(\theta | y)$ integri a uno.

Generalmente, l'integrale al denominatore della (2.6) non è valutabile analiticamente, soprattutto quando θ è di dimensione elevata. Quindi, spesso si ricorre a metodi approssimati. In particolare, nei metodi Monte Carlo, al fine di condurre l'inferenza sui parametri è sufficiente disporre di un metodo che permetta di generare campioni dalla distribuzione a posteriori conoscendone soltanto il nucleo non normalizzato, ossia il prodotto $p(y | \theta) \pi(\theta)$. La classe di metodi più utilizzata a questo scopo è quella dei metodi *Markov chain Monte Carlo* (MCMC), che consente di simulare campioni dalla distribuzione a posteriori costruendo una catena di Markov avente tale distribuzione come distribuzione stazionaria.

Una catena di Markov $\{\theta^{(s)}\}_{s \geq 0}$ è una sequenza di variabili aleatorie dipendenti tale che la distribuzione di $\theta^{(s+1)}$, condizionatamente a $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(s)}$, dipenda soltanto dal valore corrente $\theta^{(s)}$. La catena è ergodica se è irriducibile, aperiodica e ricorrente (Gelman et al., 2014, paragrafo 11.2). Sotto queste condizioni la distribuzione stazionaria è anche la distribuzione limite della catena e vale il *teorema ergodico*, per cui, data

una funzione h tale che $\mathbb{E}_{\pi(\cdot|y)}[|h(\theta)|] < \infty$,

$$\frac{1}{S} \sum_{s=1}^S h(\theta^{(s)}) \xrightarrow[S \rightarrow \infty]{\text{q.c.}} \mathbb{E}_{\pi(\cdot|y)}[h(\theta)],$$

che estende la legge dei grandi numeri al caso di campioni dipendenti generati da una catena di Markov, e giustifica l'uso delle medie campionarie sulla catena come stime delle quantità a posteriori di interesse, inclusi quantili e intervalli di credibilità ottenuti dalla distribuzione empirica dei campioni generati (Robert & Casella, 2010, paragrafo 6.2). Tra i principali algoritmi MCMC rientrano il *Gibbs sampler* e l'algoritmo di Metropolis-Hastings.

L'algoritmo di Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970) è il metodo più generale per simulare valori da una distribuzione a posteriori. Data una distribuzione di proposta $q(\cdot | \theta^{(s)})$, ad ogni iterazione si genera un valore candidato $\theta' \sim q(\cdot | \theta^{(s)})$ e lo si accetta, ponendo $\theta^{(s+1)} = \theta'$, con probabilità

$$\rho(\theta^{(s)}, \theta') = \min \left\{ 1, \frac{\pi(\theta' | y)}{\pi(\theta^{(s)} | y)} \frac{q(\theta^{(s)} | \theta')}{q(\theta' | \theta^{(s)})} \right\}; \quad (2.7)$$

in caso di rifiuto la catena resta nel valore corrente, $\theta^{(s+1)} = \theta^{(s)}$. Si noti che $\pi(\theta | y)$ compare nella (2.7) soltanto attraverso il rapporto fra le sue valutazioni in θ' e in $\theta^{(s)}$. La costante di normalizzazione della (2.6) si cancella dunque in tale rapporto, per cui il suo calcolo non è necessario. Dal punto di vista pratico, la scelta della distribuzione di proposta $q(\cdot | \theta^{(s)})$ incide sia sulla probabilità di accettazione che sulla velocità con cui la catena esplora lo spazio dei parametri, risultando quindi cruciale per l'efficienza dell'algoritmo.

Il *Gibbs sampler* (Geman & Geman, 1984; Gelfand & Smith, 1990), adottato in questa tesi per stimare i modelli proposti, si può vedere come un caso particolare dell'algoritmo di Metropolis-Hastings in cui si sfrutta una partizione del vettore dei parametri in blocchi, $\theta = (\theta_1, \dots, \theta_p)$, aggiornati ciclicamente uno alla volta. Assumendo come densità di proposta per il blocco θ_j la corrispondente *full conditional* $\pi(\theta_j | \theta_{-j}, y)$, ossia la distribuzione di θ_j condizionata a tutti gli altri blocchi e ai dati (con θ_{-j} si indica il vettore θ privato del j -esimo blocco), la probabilità di accettazione (2.7) risulta pari a uno, cosicché ciascun aggiornamento si riduce a un campionamento diretto dalla *full conditional*. Una singola iterazione consiste allora nell'aggiornare in sequenza tutti i blocchi,

$$\theta_j^{(s+1)} \sim \pi \left(\theta_j \mid \theta_1^{(s+1)}, \dots, \theta_{j-1}^{(s+1)}, \theta_{j+1}^{(s)}, \dots, \theta_p^{(s)}, y \right), \quad j = 1, \dots, p,$$

impiegando per ciascuno di essi i valori più recenti degli altri. Tale schema risulta particolarmente vantaggioso quando, come nei modelli qui considerati, le *full conditional* appartengono a famiglie note e coniugate, dalle quali è possibile campionare in modo esatto senza necessità di calibrare alcuna distribuzione di proposta (Robert & Casella, 2010), come sarebbe invece necessario in un generico Metropolis–Hastings.

Un algoritmo alternativo oggi molto diffuso è l'*Hamiltonian Monte Carlo* (HMC), originariamente proposto in ambito fisico con il nome di *Monte Carlo ibrido* (Duane et al., 1987), poiché affianca alla simulazione MCMC una dinamica deterministica (Neal, 2011; Betancourt, 2017); per una rassegna si veda Gelman et al. (2014, paragrafo 12.4). L'idea di fondo è sostituire le proposte di un algoritmo di Metropolis–Hastings a passeggiata aleatoria, tanto meno efficienti quanto più i parametri sono correlati a posteriori, con proposte che percorrono in modo coerente le regioni ad alta densità, risultando al tempo stesso distanti dal valore corrente e accettate con probabilità elevata. A questo scopo si affianca al vettore dei parametri θ un vettore ausiliario di variabili di momento r , della stessa dimensione, indipendente da θ e distribuito come $r \sim \mathcal{N}(0, M)$, dove M è una matrice tipicamente diagonale che gioca lo stesso ruolo della massa in un sistema fisico (Gelman et al., 2014, paragrafo 12.4). La distribuzione obiettivo si può scrivere come

$$p(\theta, r \mid y) = \pi(\theta \mid y) p(r) \propto \exp\{-H(\theta, r)\}, \quad H(\theta, r) = U(\theta) + \frac{1}{2} r^\top M^{-1} r, \quad (2.8)$$

dove $U(\theta) = -\log \pi(\theta \mid y)$ è detta energia potenziale e $H(\theta, r)$ energia totale, o *hamiltoniana*, del sistema.

Poiché r è indipendente da θ , la distribuzione marginale della (2.8) rispetto a θ coincide con la distribuzione a posteriori di interesse. Risulta dunque sufficiente costruire una catena sulla coppia (θ, r) e trascurarne la seconda componente per ottenere campioni da $\pi(\theta \mid y)$. Si osservi inoltre che $U(\theta)$ è nota a meno di una costante additiva, la quale non interviene né nel gradiente né nei rapporti di densità, cosicché anche in questo caso è sufficiente conoscere il nucleo non normalizzato della posteriori.

L'analogia fisica interpreta θ come la posizione di una particella che si muove su una superficie di energia potenziale $U(\theta)$, in modo che i punti a bassa energia corrispondano alle regioni ad alta densità a posteriori.

Ogni iterazione dell'algoritmo si articola in due passi. Nel primo si campiona un nuovo momento da $r \sim \mathcal{N}(0, M)$, il che equivale ad assegnare alla particella una spinta iniziale casuale. Nel secondo si determina la traiettoria che, muovendo dallo stato

corrente (θ, r) , risolve il sistema di equazioni differenziali

$$\frac{d\theta}{dt} = M^{-1}r, \quad \frac{dr}{dt} = -\nabla_{\theta}U(\theta),$$

le quali conservano il valore di $H(\theta, r)$ e dunque, per la (2.8), quello della densità congiunta. Lungo la traiettoria l'energia potenziale $U(\theta)$ e quella cinetica si scambiano continuamente, cosicché la densità a posteriori $\pi(\theta \mid y)$ varia mentre la loro somma resta costante: la particella può così allontanarsi anche di molto dal punto di partenza, attraversando regioni a densità diversa senza per questo abbandonare quelle di interesse. Il punto raggiunto al termine della traiettoria costituisce lo stato candidato. L'alternanza fra i due passi è essenziale in quanto la dinamica, conservando l'energia totale, si muove soltanto fra stati che condividono il medesimo valore di H , mentre il rinnovo del momento ne modifica la componente cinetica e rende quindi accessibili alla traiettoria successiva stati altrimenti irraggiungibili. Poiché le equazioni della dinamica hamiltoniana non ammettono in generale soluzione esplicita, la traiettoria è approssimata numericamente tramite L passi dell'integratore *leapfrog* (Gelman et al., 2014, paragrafo 12.4),

$$\begin{aligned} r &\leftarrow r - \frac{\epsilon}{2} \nabla_{\theta}U(\theta), \\ \theta &\leftarrow \theta + \epsilon M^{-1}r, \\ r &\leftarrow r - \frac{\epsilon}{2} \nabla_{\theta}U(\theta), \end{aligned} \tag{2.9}$$

dove ϵ è l'ampiezza del passo di integrazione. La discretizzazione della (2.9) non conserva esattamente l'energia totale, e l'errore così introdotto è corretto da un passo di accettazione e rifiuto analogo alla (2.7), basato sul rapporto fra le densità congiunte nello stato proposto (θ', r') e in quello corrente (θ, r) :

$$\rho((\theta, r), (\theta', r')) = \min \left\{ 1, \frac{\pi(\theta' \mid y) p(r')}{\pi(\theta \mid y) p(r)} \right\}. \tag{2.10}$$

Nella (2.10) non compare alcun rapporto fra densità di proposta, poiché lo stato (θ', r') è ottenuto da (θ, r) in modo deterministico, e al suo posto interverrebbe lo jacobiano della trasformazione, che per l'integratore *leapfrog* ha modulo unitario (Neal, 2011; Betancourt, 2017). Viene così meno la scelta della distribuzione di proposta, che nell'algoritmo di Metropolis–Hastings costituisce l'aspetto più delicato: nell'*Hamiltonian Monte Carlo* la proposta è interamente determinata dalla dinamica. Con una calibrazione adeguata di ϵ ed L , l'algoritmo fornisce stati candidati al tempo stesso distanti e frequentemente

accettati, riducendo sensibilmente l'autocorrelazione della catena rispetto a uno schema a passeggiata aleatoria, soprattutto in presenza di forte correlazione a posteriori fra i parametri.

Poiché tale calibrazione è tutt'altro che agevole, si ricorre in pratica alla variante *No-U-Turn Sampler* (NUTS) (Hoffman & Gelman, 2014), che determina automaticamente la lunghezza di ciascuna traiettoria e adatta ϵ nella fase di inizializzazione. Essa è implementata nel software **Stan** (Carpenter et al., 2017), sul quale si basa il pacchetto **footBayes** (Egidi et al., 2025) per la stima dei modelli *goal-based* assunti come termine di confronto per i modelli proposti in questa tesi.

L'efficienza nell'esplorazione dello spazio dei parametri ha però un costo per iterazione non trascurabile, in quanto ciascuno degli L passi del *leapfrog* richiede la valutazione del gradiente della log-densità a posteriori sull'intero campione (Betancourt, 2017). Per contro, quando le *full conditional* sono coniugate, ogni iterazione del *Gibbs sampler* si riduce a una sequenza di estrazioni dirette da distribuzioni note, priva tanto del calcolo di gradienti quanto di passi di accettazione e rifiuto. Tale divario, che si riflette in tempi di stima sensibilmente diversi fra i due approcci, rappresenta uno degli aspetti a favore del modello proposto in questa tesi, come verrà discusso nel Capitolo 3.

Indipendentemente dall'algoritmo impiegato, la convergenza della catena alla distribuzione stazionaria e la qualità dei campioni generati vanno verificate empiricamente. In questa tesi si adottano le due diagnostiche di uso più comune (Vehtari et al., 2021).

La prima è il fattore di riduzione di scala potenziale $\hat{R}$ (Gelman & Rubin, 1992): si inizializzano M catene da punti sovradispersi e, scartato il periodo di *burn-in*, ciascuna di esse viene divisa in due metà, trattate come sequenze distinte al fine di rilevare anche l'eventuale non stazionarietà interna a una singola catena. Si ottengono così $m = 2M$ sequenze di lunghezza n , delle quali si confronta la variabilità fra gruppi, $B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta})^2$, con quella media entro gruppo $W = \frac{1}{m} \sum_{j=1}^m s_j^2$, dove $\bar{\theta}_j$ e s_j^2 sono media e varianza campionaria della j -esima sequenza e $\bar{\theta}$ la media complessiva. Combinando le due quantità si ottiene

$$\hat{R} = \sqrt{\frac{\frac{n-1}{n} W + \frac{1}{n} B}{W}},$$

che confronta una stima della varianza a posteriori basata su tutti i campioni disponibili con quella basata sulle singole sequenze. In regime di convergenza le due stime coincidono e $\hat{R}$ tende a 1, mentre valori sensibilmente superiori segnalano che le catene non hanno ancora raggiunto una distribuzione comune. In questa tesi si adotta la soglia $\hat{R} < 1.01$ raccomandata da Vehtari et al. (2021).

La seconda metrica diagnostica è la dimensione campionaria effettiva (*effective sample size*, ESS), che quantifica il numero di estrazioni indipendenti equivalenti agli S campioni a posteriori effettivamente generati, tenendo conto della loro autocorrelazione. Si calcola

$$\text{ESS} = \frac{S}{1 + 2 \sum_{k=1}^{\infty} \rho_k},$$

dove ρ_k è l'autocorrelazione della catena a ritardo k e la somma è nella pratica troncata al ritardo oltre il quale le autocorrelazioni stimate non risultano più distinguibili dallo zero (Vehtari et al., 2021). Poiché i campioni MCMC sono di norma positivamente correlati, si ha tipicamente $\text{ESS} < S$. Un valore di ESS basso indica che la catena esplora lentamente lo spazio dei parametri e che le stime a posteriori sono affette da un errore Monte Carlo elevato. La formula fornisce una misura complessiva dell'efficienza della catena, la quale non tiene tuttavia conto della possibilità che il grado di autocorrelazione vari a seconda della regione della distribuzione a posteriori considerata. Una catena può infatti presentare un buon *mixing* nella regione centrale e una persistenza più marcata nelle code, oppure manifestare un comportamento inverso. Al fine di tener conto di tale eterogeneità, Vehtari et al. (2021) propongono di applicare la diagnostica ESS non ai campioni originali, bensì a campioni trasformati mediante una normalizzazione basata sui ranghi. Tale trasformazione, oltre a rendere la diagnostica sensibile alle differenze locali nel grado di autocorrelazione, ne garantisce l'applicabilità anche nei casi in cui la distribuzione a posteriori non possieda media o varianza finite. Sui campioni così trasformati gli autori definiscono due distinte misure di efficienza, il *bulk*-ESS e il *tail*-ESS. Il *bulk*-ESS si ottiene applicando la diagnostica standard all'intera distribuzione dei campioni normalizzati rispetto al rango. Esso quantifica l'efficienza con cui la catena esplora la regione centrale della distribuzione a posteriori e risulta pertanto particolarmente rilevante ai fini della stima di quantità quali la media e la mediana. Il *tail*-ESS quantifica invece l'efficienza della catena nello stimare le code della distribuzione da cui dipendono gli intervalli di credibilità a posteriori.

2.4 Verosimiglianza, distribuzione a posteriori e *full conditionals*

In questo paragrafo si derivano la distribuzione a posteriori congiunta e le corrispondenti distribuzioni *full conditional* dei parametri del modello, sulla base della verosimiglianza aumentata di Albert & Chib (1993), formulata in termini della variabile latente Z_i .

Alla luce dei dati osservati e assumendo le singole partite come indipendenti, la verosimiglianza è

$$L(\beta, \alpha, \sigma_\varepsilon^2 \mid y) = \prod_{i=1}^n \left\{ \Phi\left(\frac{\gamma_{y_i+1} - \eta_i}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\gamma_{y_i+2} - \eta_i}{\sigma_\varepsilon}\right) \right\}. \quad (2.11)$$

Introducendo le variabili latenti Z_i , definite in (2.1) e legate alla risposta osservata attraverso la relazione (2.3), si ottiene la verosimiglianza aumentata (*augmented likelihood*)

$$\begin{aligned} L^*(\beta, \alpha, \sigma_\varepsilon^2 \mid z, y) &= \prod_{i=1}^n \phi(z_i; \eta_i, \sigma_\varepsilon^2) \mathbb{I}(\gamma_{y_i+1} < z_i \leq \gamma_{y_i+2}) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{(z_i - \eta_i)^2}{2\sigma_\varepsilon^2}\right\} \mathbb{I}(\gamma_{y_i+1} < z_i \leq \gamma_{y_i+2}), \end{aligned}$$

dove $\phi(\cdot; \mu, \sigma^2)$ indica la densità di una $\mathcal{N}(\mu, \sigma^2)$ e $\mathbb{I}(A)$ è la funzione indicatrice di A .

La distribuzione a posteriori congiunta dei parametri e delle variabili latenti $z = (z_1, \dots, z_n)$ è

$$\pi(\beta, \sigma^2, \sigma_\varepsilon^2, \alpha, z \mid y) \propto L^*(\beta, \alpha, \sigma_\varepsilon^2 \mid z, y) \pi(\beta \mid \sigma^2) \pi(\sigma^2) \pi(\sigma_\varepsilon^2) \pi(\alpha) \quad (2.12)$$

con

$$\begin{aligned} \pi(\beta \mid \sigma^2) &= \prod_{k=1}^K \phi(\beta_{k,1}; 0, \sigma_1^2) \prod_{t=2}^T \prod_{k=1}^K \phi(\beta_{k,t}; \beta_{k,t-1}, \sigma_t^2) \\ &\propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{k=1}^K \beta_{k,1}^2\right\} \exp\left\{-\frac{1}{2} \sum_{t=2}^T \frac{1}{\sigma_t^2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2\right\}. \end{aligned} \quad (2.13)$$

Indicando con $\equiv$ l'uguaglianza a meno di costanti additive, la distribuzione a posteriori in scala logaritmica è

$$\begin{aligned}
\log \pi(\beta, \sigma^2, \sigma_\varepsilon^2, \alpha, z \mid y) &\equiv -\frac{n}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (z_i - \eta_i)^2 + \sum_{i=1}^n \log \mathbb{I}(\gamma_{y_i+1} < z_i \leq \gamma_{y_i+2}) \\
&\quad - \frac{1}{2\sigma_1^2} \sum_{k=1}^K \beta_{k,1}^2 - \frac{1}{2} \sum_{t=2}^T \frac{1}{\sigma_t^2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2 \\
&\quad - \sum_{t=2}^T \left[(a+1) \log \sigma_t^2 + \frac{b}{\sigma_t^2} \right] - (a_\varepsilon + 1) \log \sigma_\varepsilon^2 - \frac{b_\varepsilon}{\sigma_\varepsilon^2} \\
&\quad - \frac{1}{2\sigma_\alpha^2} \sum_{t=1}^T \alpha_t^2. \tag{2.14}
\end{aligned}$$

Le distribuzioni *full conditional* dei singoli parametri si ottengono isolando, nella (2.14), i soli addendi che coinvolgono il parametro di interesse, fissando tutti gli altri parametri, e riconoscendo nella somma di tali addendi il logaritmo del nucleo di una distribuzione nota. Per alleggerire la notazione, nel condizionamento si scrive $\cdot$ per indicare sinteticamente tutti i rimanenti parametri e i dati osservati.

Derivazione della *full conditional* di Z_i . Gli unici addendi della log-posteriori che dipendono da z_i sono quelli della log-verosimiglianza aumentata

$$\log \pi(Z_i \mid \cdot) \equiv -\frac{1}{2\sigma_\varepsilon^2} (z_i - \eta_i)^2 + \log \mathbb{I}(\gamma_{y_i+1} < z_i \leq \gamma_{y_i+2}).$$

Il primo addendo è il nucleo della log-densità di una $\mathcal{N}(\eta_i, \sigma_\varepsilon^2)$; il secondo è nullo se $z_i \in (\gamma_{y_i+1}, \gamma_{y_i+2}]$ e $-\infty$ altrimenti, troncando la densità a tale intervallo. Si ha dunque

$$Z_i \mid \cdot \sim \mathcal{N}(\eta_i, \sigma_\varepsilon^2) \text{ troncata in } (\gamma_{y_i+1}, \gamma_{y_i+2}]. \tag{2.15}$$

Esplicitamente, ricordando che $\gamma_1 = -0.5$ e $\gamma_2 = 0.5$

$$Z_i \mid \cdot, y_i = -1 \sim \mathcal{N}(\eta_i, \sigma_\varepsilon^2) \text{ troncata in } (-\infty, -0.5],$$

$$Z_i \mid \cdot, y_i = 0 \sim \mathcal{N}(\eta_i, \sigma_\varepsilon^2) \text{ troncata in } (-0.5, 0.5],$$

$$Z_i \mid \cdot, y_i = 1 \sim \mathcal{N}(\eta_i, \sigma_\varepsilon^2) \text{ troncata in } (0.5, +\infty).$$

Derivazione della *full conditional* di $\beta_{k,t}$ (campionamento univariato). Dalla log-posteriori, gli addendi che dipendono da $\beta_{k,t}$ sono

$$\begin{aligned} \log \pi(\beta_{k,t} \mid \cdot) \equiv & -\frac{1}{2\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t^{(k)}} (z_i - \eta_i)^2 \\ & - \frac{1}{2\sigma_t^2} (\beta_{k,t} - \beta_{k,t-1})^2 \mathbb{I}(t \geq 2) \\ & - \frac{1}{2\sigma_{t+1}^2} (\beta_{k,t+1} - \beta_{k,t})^2 \mathbb{I}(t < T). \end{aligned}$$

I tre addendi rappresentano i contributi, rispettivamente, della verosimiglianza aumentata e delle priori dinamiche che legano $\beta_{k,t}$ ai periodi $t-1$ e $t+1$.

Si analizza ora ciascun termine singolarmente, a partire dalla verosimiglianza aumentata: per ogni partita $i \in \mathcal{I}_t^{(k)}$ si espande $(z_i - \eta_i)^2$ mettendo in evidenza $\beta_{k,t}$.

- Caso $h_i = k$ (squadra k gioca in casa):

$$\eta_i = \alpha_t + \beta_{k,t} - \beta_{a_i,t} \implies z_i - \eta_i = (z_i - \alpha_t + \beta_{a_i,t}) - \beta_{k,t}.$$

Espandendo il quadrato e isolando i termini in $\beta_{k,t}$:

$$-\frac{1}{2\sigma_\varepsilon^2} (z_i - \eta_i)^2 \equiv -\frac{1}{2\sigma_\varepsilon^2} \beta_{k,t}^2 + \frac{1}{\sigma_\varepsilon^2} \underbrace{(z_i - \alpha_t + \beta_{a_i,t})}_{=: \tilde{z}_i^{(k)}} \beta_{k,t}.$$

- Caso $a_i = k$ (squadra k gioca in trasferta):

$$\eta_i = \alpha_t + \beta_{h_i,t} - \beta_{k,t} \implies z_i - \eta_i = (z_i - \alpha_t - \beta_{h_i,t}) + \beta_{k,t}.$$

Espandendo il quadrato e isolando i termini in $\beta_{k,t}$:

$$-\frac{1}{2\sigma_\varepsilon^2} (z_i - \eta_i)^2 \equiv -\frac{1}{2\sigma_\varepsilon^2} \beta_{k,t}^2 + \frac{1}{\sigma_\varepsilon^2} \underbrace{-(z_i - \alpha_t - \beta_{h_i,t})}_{=: \tilde{z}_i^{(k)}} \beta_{k,t}.$$

Generalizzando, per entrambi i casi si definisce

$$\tilde{z}_i^{(k)} = \begin{cases} z_i - \alpha_t + \beta_{a_i,t} & \text{se } h_i = k, \\ -(z_i - \alpha_t - \beta_{h_i,t}) & \text{se } a_i = k. \end{cases}$$

Sommando su tutte le $n_{k,t}$ partite si ottiene infine

$$-\frac{1}{2\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t^{(k)}} (z_i - \eta_i)^2 \equiv -\frac{n_{k,t}}{2\sigma_\varepsilon^2} \beta_{k,t}^2 + \frac{\beta_{k,t}}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t^{(k)}} \tilde{z}_i^{(k)}. \quad (2.16)$$

Si definisce $\psi_t = 1/\sigma_t^2$ per semplicità. Espandendo il quadrato, il secondo termine (presente per $t \geq 2$) si riscrive isolando i termini in $\beta_{k,t}$, ottenendo

$$\begin{aligned} -\frac{1}{2\sigma_t^2} (\beta_{k,t} - \beta_{k,t-1})^2 &= -\frac{\psi_t}{2} (\beta_{k,t}^2 - 2\beta_{k,t}\beta_{k,t-1} + \beta_{k,t-1}^2) \\ &\equiv -\frac{\psi_t}{2} \beta_{k,t}^2 + \psi_t \beta_{k,t-1} \beta_{k,t}. \end{aligned} \quad (2.17)$$

Procedendo in modo analogo per il terzo termine (presente per $t < T$), si ha

$$\begin{aligned} -\frac{1}{2\sigma_{t+1}^2} (\beta_{k,t+1} - \beta_{k,t})^2 &= -\frac{\psi_{t+1}}{2} (\beta_{k,t+1}^2 - 2\beta_{k,t+1}\beta_{k,t} + \beta_{k,t}^2) \\ &\equiv -\frac{\psi_{t+1}}{2} \beta_{k,t}^2 + \psi_{t+1} \beta_{k,t+1} \beta_{k,t}. \end{aligned} \quad (2.18)$$

Sommando le (2.16), (2.17) e (2.18), si ha

$$\begin{aligned} \log \pi(\beta_{k,t} \mid \cdot) &\equiv -\frac{1}{2} \underbrace{\left(\frac{n_{k,t}}{\sigma_\varepsilon^2} + \psi_t \mathbb{I}(t \geq 2) + \psi_{t+1} \mathbb{I}(t < T) \right)}_{=: (V_{k,t}^*)^{-1}} \beta_{k,t}^2 \\ &\quad + \underbrace{\left(\frac{1}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t^{(k)}} \tilde{z}_i^{(k)} + \psi_t \beta_{k,t-1} \mathbb{I}(t \geq 2) + \psi_{t+1} \beta_{k,t+1} \mathbb{I}(t < T) \right)}_{=: (V_{k,t}^*)^{-1} \mu_{k,t}^*} \beta_{k,t}. \end{aligned} \quad (2.19)$$

Un'espressione della forma $-\frac{1}{2}A\beta^2 + B\beta$, con $A > 0$, si può riscrivere mediante il completamento del quadrato:

$$-\frac{A}{2}\beta^2 + B\beta = -\frac{A}{2} \left(\beta - \frac{B}{A} \right)^2 + \frac{B^2}{2A}.$$

Il termine $B^2/(2A)$ non dipende da β e confluisce nella costante di proporzionalità. Si riconosce quindi in (2.19) il nucleo della log-densità di una $\mathcal{N}(\mu_{k,t}^*, V_{k,t}^*)$, ossia

$$\beta_{k,t} \mid \cdot \sim \mathcal{N}(\mu_{k,t}^*, V_{k,t}^*),$$

con

$$V_{k,t}^* = \left(\frac{n_{k,t}}{\sigma_\varepsilon^2} + \psi_t \mathbb{I}(t \geq 2) + \psi_{t+1} \mathbb{I}(t < T) \right)^{-1},$$

$$\mu_{k,t}^* = V_{k,t}^* \left(\frac{1}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t^{(k)}} \tilde{z}_i^{(k)} + \psi_t \beta_{k,t-1} \mathbb{I}(t \geq 2) + \psi_{t+1} \beta_{k,t+1} \mathbb{I}(t < T) \right).$$

Casi limite. Caso $t = 1$: non esiste il termine per $t - 1$. Ricordando che $\beta_{k,1} \sim \mathcal{N}(0, \sigma_1^2)$ (paragrafo 2.2), il suo contributo alla log-posteriori è $-\frac{1}{2\sigma_1^2} \beta_{k,1}^2$. Si ha quindi

$$V_{k,1}^* = \left(\frac{n_{k,1}}{\sigma_\varepsilon^2} + \frac{1}{\sigma_1^2} + \psi_2 \right)^{-1},$$

$$\mu_{k,1}^* = V_{k,1}^* \left(\frac{1}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_1^{(k)}} \tilde{z}_i^{(k)} + \psi_2 \beta_{k,2} \right).$$

Caso $t = T$: non esiste il termine per $t + 1$, per cui si ottiene

$$V_{k,T}^* = \left(\frac{n_{k,T}}{\sigma_\varepsilon^2} + \psi_T \right)^{-1},$$

$$\mu_{k,T}^* = V_{k,T}^* \left(\frac{1}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_T^{(k)}} \tilde{z}_i^{(k)} + \psi_T \beta_{k,T-1} \right).$$

Derivazione della *full conditional* di β_t (campionamento a blocchi). Con t fissato, si può campionare in un unico passo dalla *full conditional* dell'intero vettore $\beta_t = (\beta_{1,t}, \dots, \beta_{K,t})^\top$, anziché una componente alla volta. Condizionatamente al resto dei parametri, $\beta_t \mid \cdot \sim \mathcal{N}(\mu_t^*, V_t^*)$ è essa stessa una gaussiana K -dimensionale e, come si vedrà a breve, la sua precisione $(V_t^*)^{-1}$ ha tutti gli elementi fuori diagonale non positivi. Quando la distribuzione gaussiana da campionare ha questa proprietà, Roberts & Sahu (1997) dimostrano che il campionamento a blocchi ha un tasso di convergenza preferibile a quello di uno schema componente per componente.

Dalla log-posteriori, gli addendi che dipendono da β_t sono gli stessi della versione univariata, estesi a tutte le squadre: rappresentano i contributi, rispettivamente, della verosimiglianza aumentata su tutte le partite $i \in \mathcal{I}_t$ del periodo e delle priori dinamiche

che legano β_t ai periodi $t - 1$ e $t + 1$. A meno di costanti additive,

$$\begin{aligned} \log \pi(\beta_t \mid \cdot) &\equiv -\frac{1}{2\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t} (z_i - \eta_i)^2 \\ &\quad - \frac{1}{2\sigma_t^2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2 \mathbb{I}(t \geq 2) \\ &\quad - \frac{1}{2\sigma_{t+1}^2} \sum_{k=1}^K (\beta_{k,t+1} - \beta_{k,t})^2 \mathbb{I}(t < T). \end{aligned}$$

Si analizza ora ciascun termine singolarmente, a partire dalla verosimiglianza aumentata: per ciascuna partita $i \in \mathcal{I}_t$ si ha $\eta_i = \alpha_t + \beta_{h_i,t} - \beta_{a_i,t}$, per cui $z_i - \eta_i = (z_i - \alpha_t) - (\beta_{h_i,t} - \beta_{a_i,t})$.

Si introduce inoltre la matrice del disegno $X_t \in \mathbb{R}^{n_t \times K}$ definita riga per riga come

$$X_t[i, k] = \mathbb{I}(h_i = k) - \mathbb{I}(a_i = k), \quad i \in \mathcal{I}_t, \quad k = 1, \dots, K,$$

e il vettore delle pseudo-osservazioni $\tilde{Z}_t \in \mathbb{R}^{n_t}$ con componenti $\tilde{z}_i = z_i - \alpha_t$ per $i \in \mathcal{I}_t$. Ogni riga di X_t ha esattamente due entrate non nulle rispettivamente uguali a $+1$ in posizione della squadra di casa e a -1 in posizione della squadra ospite, per cui $(X_t \beta_t)_i = \beta_{h_i,t} - \beta_{a_i,t}$. Si ha quindi, in forma matriciale

$$\sum_{i \in \mathcal{I}_t} (z_i - \eta_i)^2 = \|\tilde{Z}_t - X_t \beta_t\|^2.$$

Espandendo il quadrato e isolando i termini in β_t si ha

$$\begin{aligned} -\frac{1}{2\sigma_\varepsilon^2} \|\tilde{Z}_t - X_t \beta_t\|^2 &= -\frac{1}{2\sigma_\varepsilon^2} (\tilde{Z}_t^\top \tilde{Z}_t - 2\tilde{Z}_t^\top X_t \beta_t + \beta_t^\top X_t^\top X_t \beta_t) \\ &\equiv -\frac{1}{2} \beta_t^\top \left(\frac{X_t^\top X_t}{\sigma_\varepsilon^2} \right) \beta_t + \beta_t^\top \left(\frac{X_t^\top \tilde{Z}_t}{\sigma_\varepsilon^2} \right). \end{aligned} \quad (2.20)$$

È utile osservare la struttura di $X_t^\top X_t \in \mathbb{R}^{K \times K}$: l'elemento diagonale (k, k) è pari a $n_{k,t}$, ossia il numero di partite giocate dalla squadra k nel periodo t , mentre l'elemento (k, k') con $k \neq k'$ è pari a $-n_{k,k',t}$, dove $n_{k,k',t}$ è il numero di scontri diretti tra k e k' nel periodo t . Espandendo il quadrato, il secondo termine (presente per $t \geq 2$) si riscrive

isolando i termini in β_t , con I_K matrice identità di ordine K :

$$\begin{aligned}
-\frac{\psi_t}{2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2 &= -\frac{\psi_t}{2} \|\beta_t - \beta_{t-1}\|^2 \\
&= -\frac{\psi_t}{2} (\beta_t^\top \beta_t - 2\beta_t^\top \beta_{t-1} + \beta_{t-1}^\top \beta_{t-1}) \\
&\equiv -\frac{1}{2} \beta_t^\top (\psi_t I_K) \beta_t + \beta_t^\top (\psi_t \beta_{t-1}). \tag{2.21}
\end{aligned}$$

Procedendo in modo analogo, il terzo termine (presente per $t < T$) si riscrive come

$$\begin{aligned}
-\frac{\psi_{t+1}}{2} \sum_{k=1}^K (\beta_{k,t+1} - \beta_{k,t})^2 &= -\frac{\psi_{t+1}}{2} \|\beta_{t+1} - \beta_t\|^2 \\
&\equiv -\frac{1}{2} \beta_t^\top (\psi_{t+1} I_K) \beta_t + \beta_t^\top (\psi_{t+1} \beta_{t+1}). \tag{2.22}
\end{aligned}$$

Sommando le (2.20), (2.21) e (2.22), a meno di costanti additive, si ha

$$\begin{aligned}
\log \pi(\beta_t \mid \cdot) &= -\frac{1}{2} \beta_t^\top \underbrace{\left(\frac{X_t^\top X_t}{\sigma_\varepsilon^2} + \psi_t I_K \mathbb{I}(t \geq 2) + \psi_{t+1} I_K \mathbb{I}(t < T) \right)}_{=: (V_t^*)^{-1}} \beta_t \\
&\quad + \underbrace{\beta_t^\top \left(\frac{X_t^\top \tilde{Z}_t}{\sigma_\varepsilon^2} + \psi_t \beta_{t-1} \mathbb{I}(t \geq 2) + \psi_{t+1} \beta_{t+1} \mathbb{I}(t < T) \right)}_{=: (V_t^*)^{-1} \mu_t^*}. \tag{2.23}
\end{aligned}$$

Un'espressione della forma $-\frac{1}{2} \beta^\top A \beta + \beta^\top b$, con A simmetrica definita positiva, si può riscrivere, in modo analogo al caso scalare, completando il quadrato

$$-\frac{1}{2} \beta^\top A \beta + \beta^\top b = -\frac{1}{2} (\beta - A^{-1}b)^\top A (\beta - A^{-1}b) + \frac{1}{2} b^\top A^{-1}b.$$

Il termine $\frac{1}{2} b^\top A^{-1}b$ è costante in β e confluisce nella costante di proporzionalità. Si riconosce quindi in (2.23) il nucleo della log-densità di una $\mathcal{N}_K(\mu_t^*, V_t^*)$. In particolare, si ha

$$\beta_t \mid \cdot \sim \mathcal{N}_K(\mu_t^*, V_t^*), \tag{2.24}$$

con

$$(V_t^*)^{-1} = \frac{X_t^\top X_t}{\sigma_\varepsilon^2} + \psi_t I_K \mathbb{I}(t \geq 2) + \psi_{t+1} I_K \mathbb{I}(t < T), \tag{2.25}$$

$$\mu_t^* = V_t^* \left(\frac{X_t^\top \tilde{Z}_t}{\sigma_\varepsilon^2} + \psi_t \beta_{t-1} \mathbb{I}(t \geq 2) + \psi_{t+1} \beta_{t+1} \mathbb{I}(t < T) \right). \tag{2.26}$$

Casi limite. Caso $t = 1$: non esiste il termine per $t - 1$, analogamente al caso scalare si ottiene

$$(V_1^*)^{-1} = \frac{X_1^\top X_1}{\sigma_\varepsilon^2} + \frac{1}{\sigma_1^2} I_K + \psi_2 I_K,$$

$$\mu_1^* = V_1^* \left(\frac{X_1^\top \tilde{Z}_1}{\sigma_\varepsilon^2} + \psi_2 \beta_2 \right).$$

Caso $t = T$: non esiste il termine per $t + 1$, quindi si ha

$$(V_T^*)^{-1} = \frac{X_T^\top X_T}{\sigma_\varepsilon^2} + \psi_T I_K,$$

$$\mu_T^* = V_T^* \left(\frac{X_T^\top \tilde{Z}_T}{\sigma_\varepsilon^2} + \psi_T \beta_{T-1} \right).$$

Dal punto di vista computazionale non conviene calcolare esplicitamente l'inversa della matrice di precisione $(V_t^*)^{-1}$. Si ricorre invece alla fattorizzazione di Cholesky

$$(V_t^*)^{-1} = LL^\top,$$

con L matrice triangolare inferiore. Indicando con q_t il secondo fattore nella (2.26) (ovvero $\mu_t^* = V_t^* q_t$), la media a posteriori si ottiene senza calcolare $(V_t^*)^{-1}$ risolvendo in sequenza i due sistemi triangolari

$$Lu = q_t, \quad L^\top \mu_t^* = u.$$

Si possono generare valori dalla *full conditional* di β_t simulando da $w \sim \mathcal{N}_K(\underline{0}, I_K)$ e calcolando

$$\beta_t = \mu_t^* + (L^\top)^{-1} w.$$

Infatti $\text{Var}(L^{-\top} w) = (L^\top)^{-1} L^{-1} = (LL^\top)^{-1} = V_t^*$, per cui $\beta_t \sim \mathcal{N}_K(\mu_t^*, V_t^*)$. Questo approccio è numericamente più stabile ed efficiente rispetto al calcolo esplicito dell'inversa (Higham, 2002; Rue & Held, 2005). Con K dell'ordine di 20, tipico di un campionato europeo, il costo computazionale è notevolmente ridotto. In Tabella A.1 è riportato il confronto tra i due metodi di campionamento proposti. Alla luce della maggiore efficienza del campionamento a blocchi, quest'ultimo sarà adottato nella fase di analisi.

Derivazione della *full conditional* di σ_ε^2 . Gli addendi della log-posteriori che dipendono da σ_ε^2 sono la verosimiglianza aumentata e la priori $\text{IGa}(\sigma_\varepsilon^2; a_\varepsilon, b_\varepsilon)$:

$$\log \pi(\sigma_\varepsilon^2 \mid \cdot) \equiv -\frac{n}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (z_i - \eta_i)^2 - (a_\varepsilon + 1) \log \sigma_\varepsilon^2 - \frac{b_\varepsilon}{\sigma_\varepsilon^2}. \quad (2.27)$$

Raccogliendo i coefficienti di $\log \sigma_\varepsilon^2$ e $1/\sigma_\varepsilon^2$ si riconosce nuovamente la densità di una gamma inversa in scala logaritmica

$$\sigma_\varepsilon^2 \mid \cdot \sim \text{IGa}\left(a_\varepsilon + \frac{n}{2}, b_\varepsilon + \frac{1}{2} \sum_{i=1}^n (z_i - \eta_i)^2\right). \quad (2.28)$$

Derivazione della *full conditional* di α_t . Gli addendi della log-posteriori che dipendono da α_t sono

$$\log \pi(\alpha_t \mid \cdot) \equiv -\frac{1}{2\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t} (z_i - \alpha_t - \beta_{h_i,t} + \beta_{a_i,t})^2 - \frac{\alpha_t^2}{2\sigma_\alpha^2}.$$

Si introduce per la partita $i \in \mathcal{I}_t$

$$r_i = z_i - \beta_{h_i,t} + \beta_{a_i,t},$$

in modo che $z_i - \eta_i = r_i - \alpha_t$. Si sviluppa l'espressione come

$$-\frac{1}{2\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t} (r_i - \alpha_t)^2 \equiv -\frac{n_t}{2\sigma_\varepsilon^2} \alpha_t^2 + \frac{\alpha_t}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t} r_i.$$

Aggiungendo il contributo della a priori

$$\log \pi(\alpha_t \mid \cdot) \equiv -\frac{1}{2} \underbrace{\left(\frac{n_t}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\alpha^2}\right)}_{=: 1/V_{\alpha_t}^*} \alpha_t^2 + \underbrace{\frac{1}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t} r_i}_{=: \mu_{\alpha_t}^*/V_{\alpha_t}^*} \alpha_t.$$

Seguendo lo stesso procedimento di completamento del quadrato visto in precedenza, la *full conditional* per α_t è

$$\alpha_t \mid \cdot \sim \mathcal{N}(\mu_{\alpha_t}^*, V_{\alpha_t}^*), \quad t = 1, \dots, T, \quad (2.29)$$

con

$$V_{\alpha_t}^* = \left(\frac{n_t}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\alpha^2}\right)^{-1}, \quad \mu_{\alpha_t}^* = \frac{V_{\alpha_t}^*}{\sigma_\varepsilon^2} \sum_{i \in \mathcal{I}_t} (z_i - \beta_{h_i,t} + \beta_{a_i,t}).$$

Derivazione della *full conditional* di σ_t^2 . Gli addendi della log-posteriori che dipendono da σ_t^2 (per $t \geq 2$) sono

$$\log \pi(\sigma_t^2 \mid \cdot) \equiv -\frac{K-1}{2} \log \sigma_t^2 - \frac{1}{2\sigma_t^2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2 - (a+1) \log \sigma_t^2 - \frac{b}{\sigma_t^2}.$$

Considerando il vincolo di identificabilità si ha $\sum_{k=1}^K \beta_{k,t} = 0$ per ogni t . Gli incrementi $\delta_{k,t} = \beta_{k,t} - \beta_{k,t-1}$ soddisfano $\sum_{k=1}^K \delta_{k,t} = 0$ e sono dunque linearmente dipendenti, appartenendo a un sottospazio di dimensione $K-1$. La distribuzione a priori di β_t , vincolata a tale sottospazio, ha densità proporzionale a $(\sigma_t^2)^{-(K-1)/2} \exp\{-\frac{1}{2\sigma_t^2} \sum_k \delta_{k,t}^2\}$. Il contributo relativo a $\log \sigma_t^2$ è dunque $-\frac{K-1}{2} \log \sigma_t^2$ e non $-\frac{K}{2} \log \sigma_t^2$. Raccogliendo i coefficienti di $\log \sigma_t^2$ e $1/\sigma_t^2$ si ha

$$\log \pi(\sigma_t^2 \mid \cdot) \equiv -\left(a + \frac{K-1}{2} + 1\right) \log \sigma_t^2 - \frac{1}{\sigma_t^2} \left(b + \frac{1}{2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2\right).$$

Si riconosce dunque la log-densità di una $\text{IGa}(a^*, b^*)$, con $a^* = a + (K-1)/2$ e $b^* = b + \frac{1}{2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2$:

$$\sigma_t^2 \mid \cdot \sim \text{IGa}\left(a + \frac{K-1}{2}, b + \frac{1}{2} \sum_{k=1}^K (\beta_{k,t} - \beta_{k,t-1})^2\right), \quad t = 2, \dots, T. \quad (2.30)$$

Derivazione delle *full conditional* di $\sigma_{k,t}^2$, $\xi_{k,t}$ e p_k (priori *slope-and-slab*). Si richiami la mistura (2.5) introdotta nel paragrafo 2.2. Per ogni squadra k e ogni periodo $t \geq 2$,

$$\sigma_{k,t}^2 \sim p_k \text{IGa}(a_{\text{sp}}, b_{\text{sp}}) + (1 - p_k) \text{IGa}(a_{\text{sl}}, b_{\text{sl}}).$$

Per campionare dalla mistura si ricorre alla stessa tecnica di *data augmentation* già impiegata per la variabile latente Z_i (Diebolt & Robert, 1994). Si introduce, per ogni coppia (k, t) , un'indicatrice discreta $\xi_{k,t} \in \{0, 1\}$ che seleziona la componente di $\sigma_{k,t}^2$,

$$\xi_{k,t} \sim \text{Bernoulli}(1 - p_k), \quad \sigma_{k,t}^2 \mid \xi_{k,t} \sim \begin{cases} \text{IGa}(a_{\text{sp}}, b_{\text{sp}}) & \text{se } \xi_{k,t} = 0 \quad (\text{slope}), \\ \text{IGa}(a_{\text{sl}}, b_{\text{sl}}) & \text{se } \xi_{k,t} = 1 \quad (\text{slab}). \end{cases} \quad (2.31)$$

Marginalizzando $\xi_{k,t}$ si recupera esattamente la (2.5), con $P(\xi_{k,t} = 0) = p_k$ e $P(\xi_{k,t} = 1) = 1 - p_k$. Il vantaggio di questa procedura è che, condizionatamente a $\xi_{k,t}$, la priori su $\sigma_{k,t}^2$ e la sua *full conditional* risultano essere distribuzioni gamma inversa in forma

chiusa. Si assegna inoltre a p_k una priori beta, con iperparametri c e $d > 0$ fissati, ossia $p_k \sim \text{Be}(c, d)$.

La distribuzione a posteriori congiunta risulta quindi

$$\pi(\beta, \sigma^2, \xi, p, \sigma_\varepsilon^2, \alpha, z \mid y) \propto L^*(\beta, \alpha, \sigma_\varepsilon^2 \mid z, y) \pi(\beta \mid \sigma^2) \pi(\sigma_\varepsilon^2) \pi(\alpha) \pi(\sigma^2 \mid \xi) \pi(\xi \mid p) \pi(p),$$

dove, indicando con $f_{\text{IG}}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} e^{-b/x}$ la densità di una $\text{IGa}(a, b)$,

$$\pi(\sigma^2 \mid \xi) = \prod_{k=1}^K \prod_{t=2}^T f_{\text{IG}}(\sigma_{k,t}^2; a_{\text{sp}}, b_{\text{sp}})^{1-\xi_{k,t}} f_{\text{IG}}(\sigma_{k,t}^2; a_{\text{sl}}, b_{\text{sl}})^{\xi_{k,t}},$$

$$\pi(\xi \mid p) = \prod_{k=1}^K \prod_{t=2}^T p_k^{1-\xi_{k,t}} (1-p_k)^{\xi_{k,t}}, \quad \pi(p) \propto \prod_{k=1}^K p_k^{c-1} (1-p_k)^{d-1}.$$

Posto $D_{k,t} = (\beta_{k,t} - \beta_{k,t-1})^2$, e indicati con $a_{\xi_{k,t}}, b_{\xi_{k,t}}$ gli iperparametri della componente attiva,

$$\log \pi(\sigma_{k,t}^2 \mid \cdot) \equiv -\frac{1}{2} \log \sigma_{k,t}^2 - \frac{D_{k,t}}{2\sigma_{k,t}^2} - (a_{\xi_{k,t}} + 1) \log \sigma_{k,t}^2 - \frac{b_{\xi_{k,t}}}{\sigma_{k,t}^2},$$

dove, raccogliendo i coefficienti di $\log \sigma_{k,t}^2$ e di $1/\sigma_{k,t}^2$, si identifica il nucleo della log-densità di una distribuzione gamma inversa,

$$\sigma_{k,t}^2 \mid \xi_{k,t}, \beta \sim \text{IGa}\left(a_{\xi_{k,t}} + \frac{1}{2}, b_{\xi_{k,t}} + \frac{1}{2} D_{k,t}\right). \quad (2.32)$$

I fattori della posteriori congiunta che dipendono da $\xi_{k,t}$ sono la sua a priori (2.31) e la densità f_{IG} della componente selezionata

$$\pi(\xi_{k,t} = j \mid \cdot) \propto P(\xi_{k,t} = j) f_{\text{IG}}(\sigma_{k,t}^2; a_j, b_j), \quad j \in \{0, 1\},$$

dove $(a_0, b_0) = (a_{\text{sp}}, b_{\text{sp}})$ e $(a_1, b_1) = (a_{\text{sl}}, b_{\text{sl}})$. La *full conditional* per $\xi_{k,t}$ è dunque una distribuzione Bernoulli, con

$$P(\xi_{k,t} = 1 \mid \sigma_{k,t}^2) = \frac{(1-p_k) f_{\text{IG}}(\sigma_{k,t}^2; a_{\text{sl}}, b_{\text{sl}})}{p_k f_{\text{IG}}(\sigma_{k,t}^2; a_{\text{sp}}, b_{\text{sp}}) + (1-p_k) f_{\text{IG}}(\sigma_{k,t}^2; a_{\text{sl}}, b_{\text{sl}})}. \quad (2.33)$$

Questa probabilità rappresenta la quota di densità, pesata per il rispettivo peso a priori, attribuibile alla componente *slab* sul totale delle due componenti. Valori elevati di $\sigma_{k,t}^2$, tipici della componente *slab* e poco plausibili sotto lo *spike*, inducono una probabilità $P(\xi_{k,t} = 1 \mid \sigma_{k,t}^2)$ prossima a 1, e viceversa. Il peso p_k regola quanto $\sigma_{k,t}^2$ debba risultare più plausibile per una componente rispetto all'altra, affinché questa risulti favorita a

posteriori. La specificazione di p_k , in linea con la proposta di Macri-Demartino et al. (2026), permette un grado di *borrowing of information* dal periodo precedente specifico per ogni compagine, distinguendo così squadre storicamente più stabili da squadre più instabili in termini di prestazioni.

Ponendo $n_{sp,k} = \#\{t \geq 2 : \xi_{k,t} = 0\}$ e $n_{sl,k} = \#\{t \geq 2 : \xi_{k,t} = 1\}$, gli addendi della log-posteriori che dipendono da p_k sono

$$\begin{aligned} \log \pi(p_k \mid \xi_k) &\equiv n_{sp,k} \log p_k + n_{sl,k} \log(1 - p_k) \\ &\quad + (c - 1) \log p_k + (d - 1) \log(1 - p_k). \end{aligned}$$

Raccogliendo i coefficienti di $\log p_k$ e $\log(1 - p_k)$ si riconosce il nucleo della densità in scala logaritmica di una distribuzione beta

$$p_k \mid \xi_k \sim \text{Be}(c + n_{sp,k}, d + n_{sl,k}). \quad (2.34)$$

L'unica modifica alle *full conditional* già derivate rispetto alla specificazione a varianza di innovazione comune riguarda il campionamento a blocchi di β_t . Nella (2.25)–(2.26) i termini $\psi_t I_K$ e $\psi_{t+1} I_K$ vanno sostituiti rispettivamente con $\Psi_t = \text{diag}(1/\sigma_{1,t}^2, \dots, 1/\sigma_{K,t}^2)$ e Ψ_{t+1} , e i termini lineari $\psi_t \beta_{t-1}$, $\psi_{t+1} \beta_{t+1}$ con $\Psi_t \beta_{t-1}$, $\Psi_{t+1} \beta_{t+1}$.

Schema collassato per l'aggiornamento di $\xi_{k,t}$. Nella *full conditional* (2.33), l'indicatore $\xi_{k,t}$ viene aggiornato condizionatamente al valore corrente della varianza di evoluzione $\sigma_{k,t}^2$. Poiché $\xi_{k,t}$ determina la componente della mistura da cui $\sigma_{k,t}^2$ proviene, le due variabili risultano fortemente dipendenti a posteriori, inducendo una forte autocorrelazione nella catena di $\xi_{k,t}$. Per ridurre tale dipendenza si adotta uno schema di *partially collapsed Gibbs sampler* (van Dyk & Park, 2008; Park & van Dyk, 2009), generalizzazione del *collapsed Gibbs sampler* di Liu (1994), nel quale il parametro $\sigma_{k,t}^2$ viene marginalizzato limitatamente all'aggiornamento di $\xi_{k,t}$. In questo modo si allevia la forte dipendenza tra i campionamenti di $\sigma_{k,t}^2$ e $\xi_{k,t}$, indotta dalla struttura della (2.33), che causa catene degeneri in un unico valore per $\xi_{k,t}$ nello schema originale (Tabella 2.1).

Fissata la coppia (k, t) e condizionatamente al valore corrente di $\beta_{k,t-1}$, la catena di dipendenza tra le distribuzioni in esame è data da

$$\begin{aligned} \xi_{k,t} &\sim \text{Bernoulli}(1 - p_k), \\ \sigma_{k,t}^2 \mid \xi_{k,t} = j &\sim \text{IGa}(a_j, b_j), \\ \beta_{k,t} \mid \beta_{k,t-1}, \sigma_{k,t}^2 &\sim \mathcal{N}(\beta_{k,t-1}, \sigma_{k,t}^2), \end{aligned}$$

con $j = 0$ componente *spike* e $j = 1$ componente *slab*. Si osserva dunque che, condizionatamente a $\beta_{k,t-1}$ e $\sigma_{k,t}^2$, $\beta_{k,t}$ è indipendente da $\xi_{k,t}$. La distribuzione congiunta, condizionata a $\beta_{k,t-1}$, si fattorizza quindi come

$$\pi(\xi_{k,t}, \sigma_{k,t}^2, \beta_{k,t} \mid \beta_{k,t-1}) = \pi(\xi_{k,t}) \pi(\sigma_{k,t}^2 \mid \xi_{k,t}) \pi(\beta_{k,t} \mid \beta_{k,t-1}, \sigma_{k,t}^2).$$

Integrando rispetto a $\sigma_{k,t}^2$ si ottiene la distribuzione congiunta marginale di $(\xi_{k,t}, \beta_{k,t})$

$$\begin{aligned} \pi(\xi_{k,t}, \beta_{k,t} \mid \beta_{k,t-1}) &= \int_0^\infty \pi(\xi_{k,t}, \sigma_{k,t}^2, \beta_{k,t} \mid \beta_{k,t-1}) d\sigma_{k,t}^2 \\ &= \pi(\xi_{k,t}) \int_0^\infty \pi(\sigma_{k,t}^2 \mid \xi_{k,t}) \pi(\beta_{k,t} \mid \beta_{k,t-1}, \sigma_{k,t}^2) d\sigma_{k,t}^2. \end{aligned}$$

Per $j \in \{0, 1\}$, si scrive quindi

$$\pi(\xi_{k,t} = j, \beta_{k,t} \mid \beta_{k,t-1}) = \pi(\xi_{k,t} = j) \int_0^\infty \pi(\sigma_{k,t}^2 \mid \xi_{k,t} = j) \pi(\beta_{k,t} \mid \beta_{k,t-1}, \sigma_{k,t}^2) d\sigma_{k,t}^2.$$

Si definisce, alleggerendo la notazione scrivendo σ^2 in luogo di $\sigma_{k,t}^2$ come variabile d'integrazione,

$$\mathcal{M}_j := \pi(\beta_{k,t} \mid \beta_{k,t-1}, \xi_{k,t} = j) = \int_0^\infty \pi(\beta_{k,t} \mid \beta_{k,t-1}, \sigma^2) \pi(\sigma^2 \mid \xi_{k,t} = j) d\sigma^2, \quad j \in \{0, 1\},$$

la distribuzione marginale di $\beta_{k,t}$ sotto la componente j , ottenuta integrando rispetto a $\sigma_{k,t}^2$. Sostituendo le densità della normale e della gamma inversa e raccogliendo le potenze di σ^2 si può scrivere

$$\mathcal{M}_j = \frac{b_j^{a_j}}{\Gamma(a_j) \sqrt{2\pi}} \int_0^\infty (\sigma^2)^{-(a_j+3/2)} \exp\left\{-\frac{B_j}{\sigma^2}\right\} d\sigma^2, \quad B_j := b_j + \frac{(\beta_{k,t} - \beta_{k,t-1})^2}{2}.$$

Con l'ausilio del cambio di variabile $u = B_j/\sigma^2$, da cui $\sigma^2 = B_j/u$ e $d\sigma^2 = -(B_j/u^2) du$, l'integrale si riconduce alla funzione Gamma di Eulero $\Gamma(\cdot)$

$$\begin{aligned} \int_0^\infty (\sigma^2)^{-(a_j+3/2)} e^{-B_j/\sigma^2} d\sigma^2 &= \int_0^\infty \left(\frac{u}{B_j}\right)^{a_j+3/2} e^{-u} \frac{B_j}{u^2} du \\ &= \frac{1}{B_j^{a_j+1/2}} \int_0^\infty u^{a_j-1/2} e^{-u} du \\ &= \frac{\Gamma(a_j + \frac{1}{2})}{B_j^{a_j+1/2}}. \end{aligned}$$

Si ottiene dunque

$$\mathcal{M}_j = \frac{\Gamma(a_j + \frac{1}{2})}{\Gamma(a_j)\sqrt{2\pi}} \cdot \frac{b_j^{a_j}}{\left[b_j + \frac{(\beta_{k,t} - \beta_{k,t-1})^2}{2} \right]^{a_j+1/2}},$$

che coincide con la densità di una t di Student con $2a_j$ gradi di libertà, parametro di posizione $\beta_{k,t-1}$ e parametro di scala $\sqrt{b_j/a_j}$,

$$\beta_{k,t} \mid \beta_{k,t-1}, \xi_{k,t} = j \sim t_{2a_j} \left(\beta_{k,t-1}, \sqrt{b_j/a_j} \right).$$

La *full conditional* di $\xi_{k,t}$ si ottiene isolando dalla log-posteriori congiunta i soli addendi che dipendono da $\xi_{k,t}$. Analogamente alla versione originale della (2.33), si ha

$$\log \pi(\xi_{k,t} = j \mid \cdot) \equiv \log P(\xi_{k,t} = j) + \log \mathcal{M}_j,$$

per cui

$$\pi(\xi_{k,t} = j \mid \cdot) \propto P(\xi_{k,t} = j) \mathcal{M}_j.$$

Dunque, in questo schema, la *full conditional* di $\xi_{k,t}$ è una Bernoulli con

$$P(\xi_{k,t} = 1 \mid \cdot) = \frac{(1 - p_k) \mathcal{M}_1}{p_k \mathcal{M}_0 + (1 - p_k) \mathcal{M}_1} = \frac{1}{1 + e^{\ell_0 - \ell_1}}, \quad (2.35)$$

con $\ell_j = \log P(\xi_{k,t} = j) + \log \mathcal{M}_j$ per $j \in \{0, 1\}$. La seconda forma è utilizzata nell'implementazione per evitare problemi di stabilità numerica associati al calcolo diretto delle densità $\mathcal{M}_0$ e $\mathcal{M}_1$. Rispetto alla (2.33), la differenza riguarda la variabile rispetto alla quale viene effettuato il confronto tra le due componenti della mistura. Nello schema originale l'evidenza a favore delle due componenti è determinata dalla densità gamma inversa valutata nel valore corrente di $\sigma_{k,t}^2$, mentre nello schema collassato dalla densità marginale di $\beta_{k,t}$ sotto la componente j . L'aggiornamento dell'indicatore non è quindi più condizionato al valore corrente della varianza, permettendo l'esplorazione delle due componenti della mistura. Lo schema collassato è adottato in tutte le stime del modello a varianza squadra-specifica presentate nel Capitolo 3. In Tabella 2.1 sono riportati i confronti tra i due schemi di campionamento in termini di *effective sample size*.

Nello schema originale, $\xi_{k,t}$ assume lo stesso valore per l'intera catena in 216 delle 224 coppie di (k, t) . In tali casi, la catena non esplora entrambe le componenti della mistura (2.31), dunque, l'elevato ESS di $\sigma_{k,t}^2$ osservato non riflette una buona esplorazione

TABELLA 2.1: Confronto tra gli schemi di campionamento in termini di *effective sample size* (ESS) per i parametri $\xi_{k,t}$ e $\sigma_{k,t}^2$. I risultati si riferiscono al modello stimato sui dati della Serie A relativi alle stagioni 2020/21–2024/25, utilizzando 4 catene MCMC da 5000 iterazioni ciascuna, di cui 1500 di *burn-in*.

Parametro	Schema	ESS media	ESS mediana
$\xi_{k,t}$	Originale	0.1	0.0
$\xi_{k,t}$	Collassato	1984.2	1932.8
$\sigma_{k,t}^2$	Originale	11891.7	12739.7
$\sigma_{k,t}^2$	Collassato	4232.5	4260.5

dell'intera distribuzione a posteriori. Come mostrato nell'esempio in Figura 2.3, lo schema collassato permette invece di esplorare entrambe le componenti della mistura, in quanto le catene MCMC relative a $\xi_{k,t}$ non risultano più degeneri in uno stesso valore, mantenendo buoni valori di ESS per entrambi i parametri.

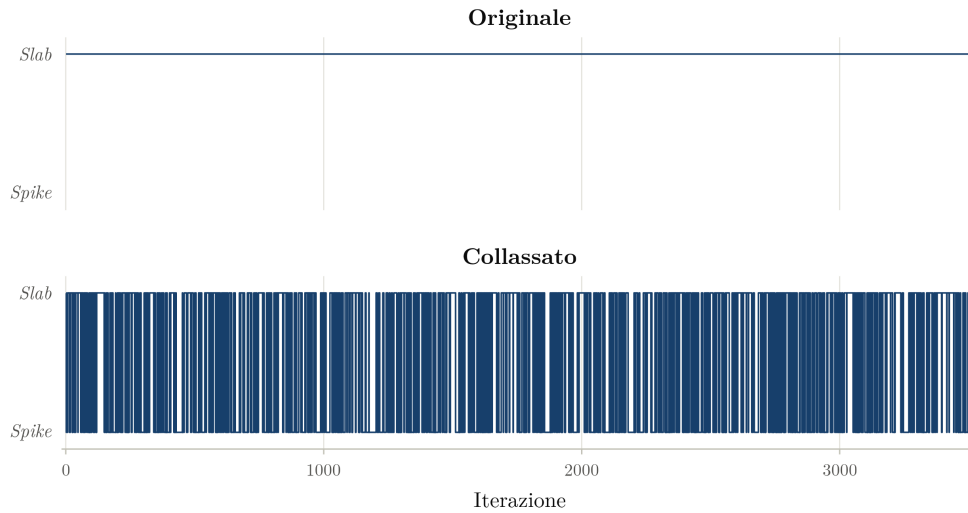


FIGURA 2.3: *Traceplot* di una catena MCMC di $\xi_{21,4}$, schema originale (sopra) e collassato (sotto). Gli iperparametri sono stati fissati a $(a_{sp}, b_{sp}) = (11, 0.1)$, $(a_{sl}, b_{sl}) = (2.5, 1)$, $(c, d) = (1, 1)$.

Si definiscono i due algoritmi di *Gibbs sampling* impiegati nel seguito, a partire dalle distribuzioni *full conditional* derivate in questo paragrafo. L'Algoritmo 1 riporta lo schema per il modello DOP e l'Algoritmo 2 quello per il modello DAOP. I due schemi differiscono unicamente nel passo relativo alle varianze di innovazione. Il vincolo di identificabilità sulle forze latenti si applica ricentrando le forze a ogni iterazione del campionario, ossia sottraendo a ciascun $\beta_{k,t}$ la media di periodo $\bar{\beta}_t = \frac{1}{K} \sum_{k=1}^K \beta_{k,t}$.

Algoritmo 1 Algoritmo di *Gibbs sampling* per il modello dinamico a varianza di innovazione comune (DOP).

```

1: Inizializzare  $\beta^{(0)}$ ,  $\alpha^{(0)}$ ,  $\sigma_\varepsilon^{2(0)}$  e  $\sigma_t^{2(0)}$ , per  $t = 2, \dots, T$ 
2: for  $s = 1, \dots, S$  do
3:   for  $i = 1, \dots, n$  do
4:     Campionare  $Z_i^{(s)}$  dalla normale troncata (2.15)
5:   end for
6:   for  $t = 1, \dots, T$  do
7:     Campionare  $\beta_t^{(s)}$  dalla normale (2.24)
8:     Ricentrare  $\beta_t^{(s)} \leftarrow \beta_t^{(s)} - \overline{\beta_t^{(s)}} \mathbf{1}_K$ 
9:   end for
10:  for  $t = 2, \dots, T$  do
11:    Campionare  $\sigma_t^{2(s)}$  dalla IGa (2.30)
12:  end for
13:  Campionare  $\sigma_\varepsilon^{2(s)}$  dalla IGa (2.28)
14:  for  $t = 1, \dots, T$  do
15:    Campionare  $\alpha_t^{(s)}$  dalla normale (2.29)
16:  end for
17: end for

```

Algoritmo 2 Algoritmo di *Gibbs sampling* per il modello dinamico a varianza specifica per squadra (DAOP).

```

1: Inizializzare  $\beta^{(0)}, \alpha^{(0)}, \sigma_\varepsilon^{2(0)}, \sigma_{k,t}^{2(0)}$  e  $p_k^{(0)}$ , per  $t = 2, \dots, T$  e  $k = 1, \dots, K$ 
2: for  $s = 1, \dots, S$  do
3:   for  $i = 1, \dots, n$  do
4:     Campionare  $Z_i^{(s)}$  dalla normale troncata (2.15)
5:   end for
6:   for  $t = 1, \dots, T$  do
7:     Campionare  $\beta_t^{(s)}$  dalla normale (2.24)
8:     Ricentrare  $\beta_t^{(s)} \leftarrow \beta_t^{(s)} - \overline{\beta_t^{(s)}} \mathbf{1}_K$ 
9:   end for
10:  for  $k = 1, \dots, K$  e  $t = 2, \dots, T$  do
11:    Campionare  $\xi_{k,t}^{(s)}$  dalla Bernoulli (2.35)
12:    Campionare  $\sigma_{k,t}^{2(s)}$  dalla IGa (spike o slab, secondo  $\xi_{k,t}^{(s)}$ ) (2.32)
13:  end for
14:  for  $k = 1, \dots, K$  do
15:    Campionare  $p_k^{(s)}$  dalla beta (2.34)
16:  end for
17:  Campionare  $\sigma_\varepsilon^{2(s)}$  dalla IGa (2.28)
18:  for  $t = 1, \dots, T$  do
19:    Campionare  $\alpha_t^{(s)}$  dalla normale (2.29)
20:  end for
21: end for

```

Capitolo 3

Applicazione

In questo capitolo le specificazioni del modello introdotte nel Capitolo 2 vengono applicate a dati reali di quattro campionati europei. Nel paragrafo 3.1 sono descritti i dati e la loro suddivisione in periodi, nel paragrafo 3.2 sono definiti gli schemi di previsione e i modelli posti a confronto, nel paragrafo 3.3 sono motivati i valori assegnati agli iperparametri e nel paragrafo 3.4 sono introdotte le metriche impiegate nella valutazione. Nel paragrafo 3.5 si analizzano e interpretano i parametri stimati dai modelli mentre nel paragrafo 3.6 vengono riportati i risultati circa la capacità predittiva dei modelli. Il paragrafo 3.7 chiude il capitolo con la diagnostica di convergenza delle catene.

Tutte le analisi sono state condotte in R, versione 4.5.2 (R Core Team, 2025).

3.1 Dati e struttura temporale

Il modello presentato nel capitolo precedente viene applicato a dati reali di quattro campionati europei di calcio: Serie A, Bundesliga, Premier League e Liga, rispettivamente le massime divisioni di Italia, Germania, Inghilterra e Spagna. I dati, relativi alle cinque stagioni dal 2020/21 al 2024/25, sono disponibili sul sito football-data.co.uk.

Ogni stagione è suddivisa in due periodi, come proposto da Macrì-Demartino et al. (2026), per un totale di $T = 10$ periodi nell'arco delle cinque stagioni considerate (Tabella 3.1). La suddivisione temporale è determinata individuando il più lungo intervallo senza incontri nella parte centrale del campionato, tra dicembre e gennaio. Per la Serie A, la Bundesliga e la Liga, il primo periodo della stagione comprende gli incontri disputati dall'inizio della stagione fino alla sosta invernale, mentre il secondo comprende quelli successivi alla ripresa del campionato fino alla sua conclusione. Nel caso della Premier League, che non prevede una sosta invernale, la separazione è fissata in corrispondenza dell'apertura della finestra di mercato di gennaio. Per la stagione 2022/23 si

considera come riferimento l'interruzione dei campionati tra la metà di novembre e la fine di dicembre, in occasione della Coppa del Mondo FIFA in Qatar. Al fine di mantenere una coerenza temporale nello schema di previsione a finestra crescente, proposto nel paragrafo 3.2, le partite di ogni stagione di ciascun campionato sono inoltre raggruppate in giornate, indicizzate da $g = 1, \dots, G$, dove G è il numero totale di giornate della stagione. Una giornata raccoglie le partite di un turno di campionato, e in essa ciascuna squadra compare al più una volta. Le giornate così definite sono ricostruite a partire dalla sola data degli incontri, procedendo in ordine cronologico e aprendo una nuova giornata non appena una squadra vi comparirebbe una seconda volta, pertanto non coincidono necessariamente con le giornate di calendario ufficiali delle rispettive leghe. Tale ricostruzione risulta robusta ai rinvii delle partite, le quali vengono attribuite ad una giornata coerente con la data in cui sono state effettivamente disputate, anziché con quella di calendario originariamente prevista.

La suddivisione temporale adottata mira a individuare periodi della stagione in cui possono verificarsi cambiamenti rilevanti nelle squadre, in grado di incidere sul loro rendimento. Le transizioni da una stagione all'altra sono caratterizzate da una pausa di diversi mesi e da cambiamenti strutturali durante la sessione di mercato estiva, risultando tipicamente più impattanti sulle prestazioni delle squadre rispetto alle transizioni entro le stagioni (Koopman & Lit, 2015). La proposta di Macrì-Demartino et al. (2026) basata sul meccanismo di commensurabilità introdotto nel paragrafo 2.2, permette di tenere conto di questa eterogeneità.

3.2 Schemi di previsione

La capacità predittiva delle specificazioni presentate nel Capitolo 2 è valutata sui dati descritti nel paragrafo 3.1. Si precisa anzitutto come dal modello stimato si ottengano le probabilità previste, e si descrivono quindi i due schemi secondo i quali esse vengono prodotte e confrontate.

Le espressioni riportate nel paragrafo 2.1 forniscono la probabilità di ciascun esito condizionatamente ai parametri. La previsione per una partita non ancora osservata si ottiene mediante la distribuzione predittiva a posteriori

$$P(Y_i = j \mid y) = \int_{\Theta} P(Y_i = j \mid \eta_i, \sigma_\varepsilon^2) \pi(\theta \mid y) d\theta, \quad j \in \{-1, 0, 1\},$$

dove θ raccoglie tutti i parametri del modello. L'integrale non è disponibile in forma chiusa. Trattandosi però del valore atteso a posteriori della funzione $h(\theta) = P(Y_i = j \mid$

TABELLA 3.1: Soglia temporale fra primo e secondo periodo di ciascuna stagione (data dell'ultima partita disputata nel primo periodo) e numero di partite incluse in ciascun periodo, per i quattro campionati considerati.

Campionato	Stagione	Soglia temporale	Primo periodo	Secondo periodo
Serie A	2020/21	23/12/2020	138	242
	2021/22	22/12/2021	189	191
	2022/23	13/11/2022	150	230
	2023/24	23/12/2023	170	210
	2024/25	23/12/2024	168	212
Bundesliga	2020/21	20/12/2020	117	189
	2021/22	19/12/2021	153	153
	2022/23	13/11/2022	135	171
	2023/24	20/12/2023	143	163
	2024/25	22/12/2024	135	171
Premier League	2020/21	04/01/2021	164	216
	2021/22	03/01/2022	191	189
	2022/23	13/11/2022	146	234
	2023/24	02/01/2024	198	182
	2024/25	06/01/2025	199	181
Liga	2020/21	12/01/2021	177	203
	2021/22	22/12/2021	181	199
	2022/23	10/11/2022	140	240
	2023/24	23/12/2023	180	200
	2024/25	22/12/2024	181	199

$\eta_i, \sigma_\varepsilon^2$), il teorema ergodico enunciato nel paragrafo 2.3 ne fornisce un'approssimazione come media sulle estrazioni della catena. Indicando con $\eta_i^{(s)}$ e $\sigma_\varepsilon^{(s)}$ i valori del predittore lineare e della deviazione standard dell'errore nella s -esima delle S estrazioni di una catena, si può stimare la probabilità di ogni esito come

$$\hat{P}(Y_i = j \mid y) = \frac{1}{S} \sum_{s=1}^S \left[\Phi\left(\frac{\gamma_{j+2} - \eta_i^{(s)}}{\sigma_\varepsilon^{(s)}}\right) - \Phi\left(\frac{\gamma_{j+1} - \eta_i^{(s)}}{\sigma_\varepsilon^{(s)}}\right) \right], \quad j \in \{-1, 0, 1\}.$$

Le previsioni sono prodotte e valutate secondo due schemi distinti. Il primo confronto oppone le specificazioni proposte (SOP, DOP e DAOP) a cinque modelli *goal-based* per dati di conteggio (Dixon & Coles (1997), Poisson con specificazione di Maher (1982), binomiale negativa, Skellam (1946) e Skellam a inflazione di zeri) ciascuno dotato della specificazione *spike-and-slab* sulla precisione di innovazione dei parametri proposta da Macrì-Demartino et al. (2026). In particolare, la valutazione avviene in tre scenari distinti che ricalcano quelli adottati dagli autori, ovvero la previsione degli esiti degli incontri della seconda metà delle stagioni 2023/24 e 2024/25, e dell'ultima giornata 2024/25. In ciascuno scenario l'insieme di stima è costituito da tutte le partite precedenti. In questo modo, il modello dispone delle informazioni osservate fino all'inizio del

periodo considerato e le previsioni vengono effettuate sugli incontri successivi. Per i primi due scenari, tale impostazione consente di valutare la capacità del modello di cogliere eventuali variazioni nel rendimento delle squadre intervenute nel corso della stagione, anche in relazione a cambiamenti quali trasferimenti, cambi di allenatore o modifiche tattiche (Macrì-Demartino et al., 2026). Il terzo scenario è invece certamente il più incerto, in quanto l'ultima giornata dei campionati riserva tradizionalmente risultati a sorpresa, oltre ad essere lo scenario con meno partite disponibili e pertanto producendo metriche meno affidabili.

Il secondo confronto riguarda esclusivamente le specificazioni dinamiche, DOP e DAOP, ed è condotto secondo uno schema di previsione a finestra crescente (*expanding window*) sulla seconda parte della stagione 2024/25 nei quattro campionati. Riservando un periodo di inizializzazione corrispondente alle prime cinque giornate del periodo T , necessario a garantire una base minima di dati per la stima del modello, la procedura ha inizio a partire dalla sesta giornata del suddetto periodo. Fissato un orizzonte di previsione $H \in \{1, \dots, 10\}$, espresso in numero di giornate di campionato, ogni modello è stimato considerando nell'insieme di stima tutte le partite disputate fino alla giornata g , e le previsioni vengono valutate sulle H giornate successive. La procedura viene ripetuta iterativamente per $g + 1$ fino a quando vale $g < G - H$. In questo modo si possono confrontare i modelli in uno scenario più realistico di utilizzo reale, valutando inoltre la capacità predittiva sia nel breve che nel lungo termine. I modelli *goal-based* non sono inclusi in questo secondo schema per ragioni di costo computazionale, dal momento che esso richiede una nuova stima a ogni origine della previsione. Il medesimo insieme di stime è invece riutilizzato per tutti gli orizzonti, poiché l'insieme di stima dipende dall'origine e non da H .

Si affiancano alle previsioni dei modelli, come termine di paragone esterno, le previsioni implicite nelle quote rilevate prima di ogni incontro, proposte dal *bookmaker* Bet365. Indicando con $q_{i,r}$ la quota associata all'esito r della partita i , il reciproco $1/q_{i,r}$ fornisce la probabilità implicita al lordo del margine dell'operatore; la somma dei tre reciproci eccede infatti l'unità. Per semplicità il margine viene rimosso per normalizzazione, ponendo $p_{i,r} = (1/q_{i,r}) / \sum_{s=1}^3 (1/q_{i,s})$, ripartendolo proporzionalmente fra i tre esiti. Un'alternativa più elaborata, in particolare la procedura proposta da Shin (1993), consiste nell'attribuire parte del margine alla presenza di scommettitori informati anziché distribuirlo in proporzione.

Tale riferimento non costituisce un modello concorrente, in quanto le quote proposte dai *bookmakers* si basano su un insieme di informazioni sul contesto di ogni incontro difficilmente eguagliabile, come ad esempio le formazioni annunciate, le condizioni dei

singoli giocatori, i movimenti di mercato e i modelli proprietari degli operatori, oltre all'informazione dalle scommesse raccolte e molto altro. Superarle non è un esito atteso né costituisce l'obiettivo del confronto, come osservano Macrì-Demartino et al. (2026). Il riferimento serve a paragonare l'accuratezza dei modelli rispetto a una fonte esterna ben calibrata, e la distanza da esso misura quanta parte dell'informazione disponibile ciascun modello riesca a recuperare dai soli risultati storici.

3.3 Iperparametri

Per svolgere le analisi, si implementa il modello bayesiano presentato nel Capitolo 2, nelle due specificazioni proposte. Le distribuzioni a priori introdotte nel paragrafo 2.2 richiedono di fissare alcuni iperparametri. I valori sono scelti in modo che le variazioni di rendimento ammesse a priori siano di ampiezza plausibile. L'analisi di sensitività riportata al termine del paragrafo mostra che la sostanza dei risultati non risulta sensibile a tali scelte.

Per scegliere valori plausibili per gli iperparametri, si può fare riferimento alla scala indotta dai valori delle soglie scelti nella specificazione del modello. Fissandole a -0.5 e 0.5 , ossia nei punti medi fra i valori -1 , 0 e 1 della differenza reti, la variabile latente Z_i si interpreta come una approssimazione continua della differenza reti dell'incontro. Il suo valore atteso $\alpha_t + \beta_{h,t} - \beta_{a,t}$ si può considerare allora come la differenza reti attesa, della quale α_t rappresenta la parte dovuta al vantaggio di giocare in casa. $\beta_{k,t}$ approssima dunque di quante reti la squadra k prevarrebbe in campo neutro su una squadra di rendimento medio.

Facendo riferimento a dati storici dei maggiori campionati europei, la differenza reti per partita si colloca genericamente entro un intervallo di circa ± 1.5 . Considerando questo intervallo si possono dunque scegliere dei valori ragionevoli per gli iperparametri sulle forze latenti.

Rispetto alla scala considerata si assegnano delle distribuzioni a priori poco informative per i parametri $\beta_{k,1}$, α e σ_ε^2 . In particolare si assegna alle forze del periodo iniziale e all'effetto casa una normale di varianza 5, che corrisponde a un intervallo di credibilità a priori al 95% di ± 4.38 . Alla varianza dell'errore si assegna una $\text{IGa}(2, 1)$. Entrambe sono distribuzioni volutamente ampie rispetto alla scala empirica appena descritta, e influiscono sulle stime in misura trascurabile.

Per quanto riguarda le varianze di evoluzione dei parametri, per il modello DOP si adotta $\sigma_t^2 \sim \text{IGa}(2, 0.1)$, caratterizzata da una media pari a 0.1, coerente con una variabilità attesa a priori dell'innovazione $\beta_{k,t} - \beta_{k,t-1}$ sulla scala considerata, con

$E[\sigma_t] \approx 0.28$, di circa un terzo di rete per partita. Fissando il parametro di forma della distribuzione a priori a 2, si assicura una media finita e una coda destra pesante, cosicché variazioni ampie restano possibili quando i dati le richiedono. Accumulandosi sui nove passaggi di periodo del campione, tale innovazione ammette una deriva sufficiente a consentire un cambiamento di posizione anche marcato ma non a percorrere la graduatoria per intero.

Nel modello DAOP la varianza di innovazione è una mistura di due componenti, che distinguono il regime ordinario da quello eccezionale. Per lo *spike* si fissa la distribuzione a priori $\text{IGa}(11, 0.1)$, con una media pari a 0.01 e varianza molto ridotta, pensata per ammettere variazioni minime, e descrive lo scenario in cui rendimento non cambia in misura apprezzabile. La *slab*, $\text{IGa}(2.5, 1)$, ammette una variazione attesa a priori di circa otto decimi nella scala della deviazione standard. Questo è coerente con un genere di variazione che si può osservare in corrispondenza di mutamenti sostanziali.

La scelta adottata in questa tesi risulta coerente con le distribuzioni a priori scelte in Macrì-Demartino et al. (2026), dove la commensurabilità fra periodi consecutivi è governata da un parametro di precisione al quale è assegnata una mistura di normali troncate in zero, $\mathcal{N}^+(9, 1.5^2)$ per la *spike* e $\mathcal{N}^+(0, 3^2)$ per la *slab*; ne discendono deviazioni standard di innovazione di mediana a priori pari a 0.11 e a 0.50, in linea con i valori di 0.10 e 0.82 adottati qui. Le distribuzioni a priori adottate non si discostano pertanto dalla pratica corrente.

Il peso della mistura non è fissato ma stimato, a partire da una distribuzione a priori $\text{Be}(c, d)$ della quale si propongono due calibrazioni nel resto delle analisi. Una priori $\text{Be}(1, 1)$, impiegata da Macrì-Demartino et al. (2026), mentre la seconda proposta è una $\text{Be}(3, 1)$, che sposta la massa verso la componente concentrata e risulta più prudente, portando il peso atteso della *spike* da 0.50 a 0.75 riducendo quindi a priori la probabilità di *shock*. Il risultato che si attende è, con la distribuzioni a priori più parsimoniosa, di ottenere delle stime intermedie tra il modello DOP e il modello DAOP dotato di priori $\text{Be}(1, 1)$.

Nell'Appendice A.3 si riportano i risultati di un'analisi di sensitività sugli iperparametri delle varianze di innovazione. Per il modello DOP le combinazioni esaminate sono ottenute incrociando tre valori del parametro di forma con tre della media a priori di σ_t^2 , mentre per il modello DAOP riscalando di un fattore 0.5, 1 e 2 le mediane a priori delle due componenti della mistura. Né l'ordinamento delle squadre né la lettura delle traiettorie ne risultano modificati.

3.4 Metriche predittive

La capacità predittiva di ogni modello è valutata tramite quattro metriche: il Brier score, il *Ranked Probability Score* (RPS), la media delle probabilità corrette (ACP) e lo pseudo- R^2 .

Si consideri un campione di verifica di n partite. Per la partita $i = 1, \dots, n$ si indichino con $p_{i,1}$, $p_{i,2}$ e $p_{i,3}$ le probabilità previste, ottenute come descritto nel paragrafo 3.2, rispettivamente per la vittoria della squadra ospite, il pareggio e la vittoria della squadra di casa, e con $y_i \in \{-1, 0, 1\}$ l'esito effettivamente osservato, nella stessa codifica di Y_i introdotta nel paragrafo 2.1. Per indicizzare le probabilità previste si pone $y_i^* = y_i + 2 \in \{1, 2, 3\}$, e si definisce $\delta_{i,r} = 1$ se $r = y_i^*$ e $\delta_{i,r} = 0$ altrimenti.

Il Brier score (Brier, 1950) è la media, sulle partite dell'insieme di verifica, degli scarti al quadrato fra il vettore delle probabilità previste e il vettore che indica l'esito osservato,

$$\text{Brier} = \frac{1}{n} \sum_{i=1}^n \sum_{r=1}^3 (p_{i,r} - \delta_{i,r})^2,$$

e assume valori compresi fra 0, in corrispondenza di previsioni perfette, e 2.

Il *Ranked Probability Score* (Epstein, 1969), adottato da Constantinou & Fenton (2012) come misura di riferimento nella previsione dei risultati calcistici, è l'unica fra le misure considerate a tenere conto dell'ordinamento naturale delle tre categorie, in quanto opera sulle probabilità cumulate e penalizza pertanto in misura maggiore gli errori più distanti dall'esito osservato. È definito come

$$\text{RPS} = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \sum_{r=1}^2 \left(\sum_{j=1}^r p_{i,j} - \sum_{j=1}^r \delta_{i,j} \right)^2,$$

assume valori tra 0 e 1 e, come per il Brier score, valori bassi indicano una migliore qualità della previsione.

La media delle probabilità corrette (*average of correct probabilities*, ACP) è la media aritmetica delle probabilità assegnate dal modello all'esito effettivamente realizzato,

$$\text{ACP} = \frac{1}{n} \sum_{i=1}^n p_{i,y_i^*}.$$

A differenza delle due misure precedenti, l'ACP cresce con la qualità della previsione in un intervallo tra 0 e 1.

Lo pseudo- R^2 (Dobson & Goddard, 2001) è costruito in modo analogo all'ACP,

sostituendo però alla media aritmetica la media geometrica,

$$\text{pseudo-}R^2 = \left(\prod_{i=1}^n p_{i,y_i^*} \right)^{1/n}.$$

Rispetto alla media aritmetica, la media geometrica penalizza più severamente le partite in cui il modello assegna una probabilità bassa all'esito poi realizzato. La metrica anche in questo caso assume valori tra 0 e 1, dove valori prossimi a 1 indicano una previsione accurata.

3.5 Analisi dei parametri stimati

3.5.1 Effetto casa

La Figura 3.1 riporta, per la Premier League, le distribuzioni a posteriori dell'effetto casa α_t per le stagioni dalla 2020/21 alla 2024/25. Nei primi due periodi, l'effetto casa risulta sostanzialmente concentrato in prossimità dello zero. A partire dai periodi successivi, invece, la distribuzione si sposta verso valori positivi, evidenziando un effetto casa più marcato a favore della squadra ospitante. Tale andamento appare coerente con la particolare condizione in cui si è disputata la stagione 2020/21, caratterizzata dall'assenza di pubblico negli stadi a causa delle restrizioni legate alla pandemia di COVID-19.

Questo risultato motiva l'utilizzo di un effetto casa specifico per ciascun periodo, ma non richiede necessariamente l'introduzione di una struttura dinamica analoga a quella adottata per le abilità delle squadre. In particolare, l'assunzione di un'evoluzione secondo un *random walk* implicherebbe che l'effetto casa di un determinato periodo sia informativo rispetto a quello del periodo precedente e che le variazioni nel tempo siano, almeno a priori, gradualmente. Tale ipotesi appare meno naturale nel caso dell'effetto casa, la cui intensità può essere influenzata da fattori esterni e difficilmente prevedibili, quali la presenza del pubblico, restrizioni allo svolgimento delle partite o altre modifiche nelle condizioni di gioco. Non vi è quindi una motivazione altrettanto forte per assumere che α_t debba evolvere in modo graduale nel tempo.

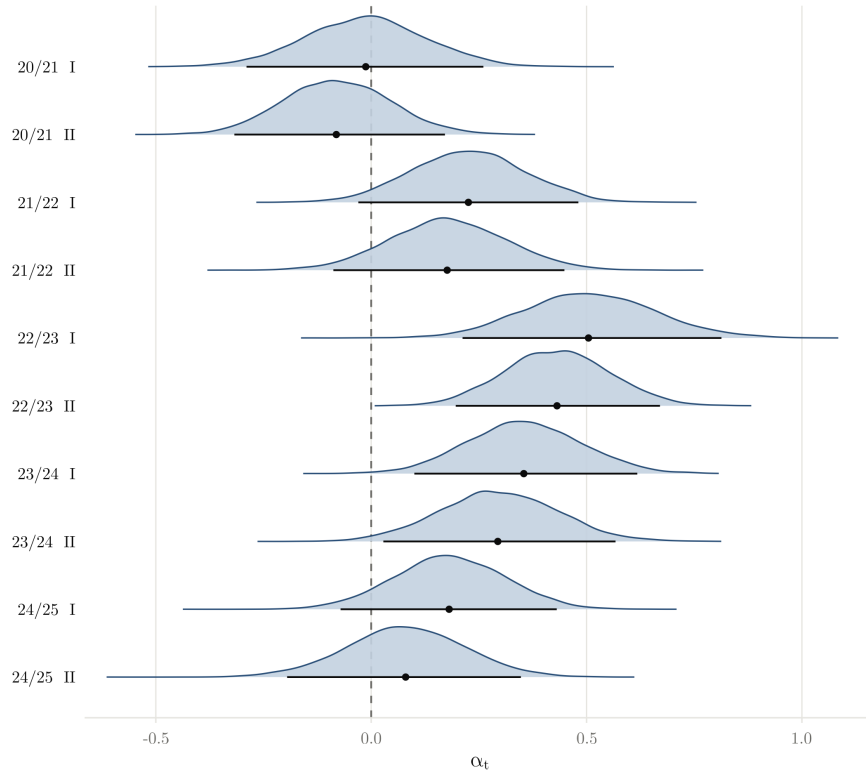


FIGURA 3.1: Distribuzione a posteriori dell'effetto casa α_t per periodo, Premier League, modello DOP. Punto e barre: mediana e intervalli di credibilità al 95%.

3.5.2 Ranking

Le forze latenti inducono, in ciascun periodo, un *ranking* delle squadre che può essere seguito nel tempo. Le Figure 3.2–3.4 riportano tale ordinamento per la Serie A sotto le tre specificazioni, con una riga per squadra e i periodi in ascissa. Le squadre compaiono soltanto nei periodi in cui hanno militato nella massima divisione, pertanto le promozioni e le retrocessioni si leggono come ingressi e uscite dal grafico. Qui e nel seguito del capitolo la forza latente di una squadra in un periodo è sintetizzata dalla mediana della sua distribuzione a posteriori, e su tale sintesi sono costruiti sia gli ordinamenti sia le variazioni riportate nel testo.

L'ordinamento è dunque indotto da una sintesi puntuale e non riflette l'incertezza che accompagna le forze, la quale non è trascurabile rispetto alle distanze che l'ordinamento riassume. L'intervallo di credibilità al 95% della forza di una singola squadra ha ampiezza mediana pari a 0.76 sotto il modello DOP e a 0.99 sotto il modello DAOP, a fronte di una distanza mediana fra la prima e l'ultima squadra del campionato pari rispettivamente a 1.89 e 2.25. Le posizioni di squadre con forze simili vanno pertanto lette come indicazioni, mentre i confronti fra squadre distanti in graduatoria sono sostenuti dai dati in modo più netto.

Osservando la Figura 3.2, che riporta il *ranking* dinamico prodotto dal modello DOP, il primo elemento che emerge è la marcata stabilità del vertice. L'Inter occupa la prima posizione in nove periodi su dieci, cedendola soltanto nella prima metà della stagione 2022/23, quando è superata dal Napoli. Altre squadre che mostrano un rendimento stabile nel tempo sono la Juventus e la Lazio, mentre esempi di compagini che variano in maniera sostanziale sono il Sassuolo, che passa dalla ottava posizione alla penultima posizione nel giro di quattro stagioni, il Monza e la Sampdoria.

Scendendo nella graduatoria il *ranking* latente individua con buona precisione le squadre destinate a lasciare la categoria. Si assumano come indicazione le tre posizioni finali di ciascuna stagione. Sotto il modello DOP e sotto il modello DAOP con priori $\text{Be}(1, 1)$, che in questo confronto forniscono le medesime terne, dieci delle dodici retrocessioni avvenute nel periodo considerato risultano correttamente identificate, con corrispondenza esatta in tre stagioni su quattro. L'eccezione è la stagione 2021/22, nella quale entrambe collocano fra le ultime tre Sampdoria e Spezia, che si salvano, mancando Cagliari e Genoa. Il modello DAOP con priori $\text{Be}(3, 1)$, che vi individua invece il Cagliari, porta il conteggio delle retrocessioni correttamente identificate a undici su dodici.

Il confronto fra le tre figure mostra ordinamenti largamente sovrapponibili. Nell'ultimo periodo, in nessun confronto tra i *rankings* proposti dalle tre specificazioni più di cinque squadre su venti occupano posizioni differenti, e lo scarto non supera mai le due posizioni. Estendendo il confronto a tutti i periodi il numero di squadre collocate diversamente arriva a dieci e lo scarto massimo a quattro posizioni.

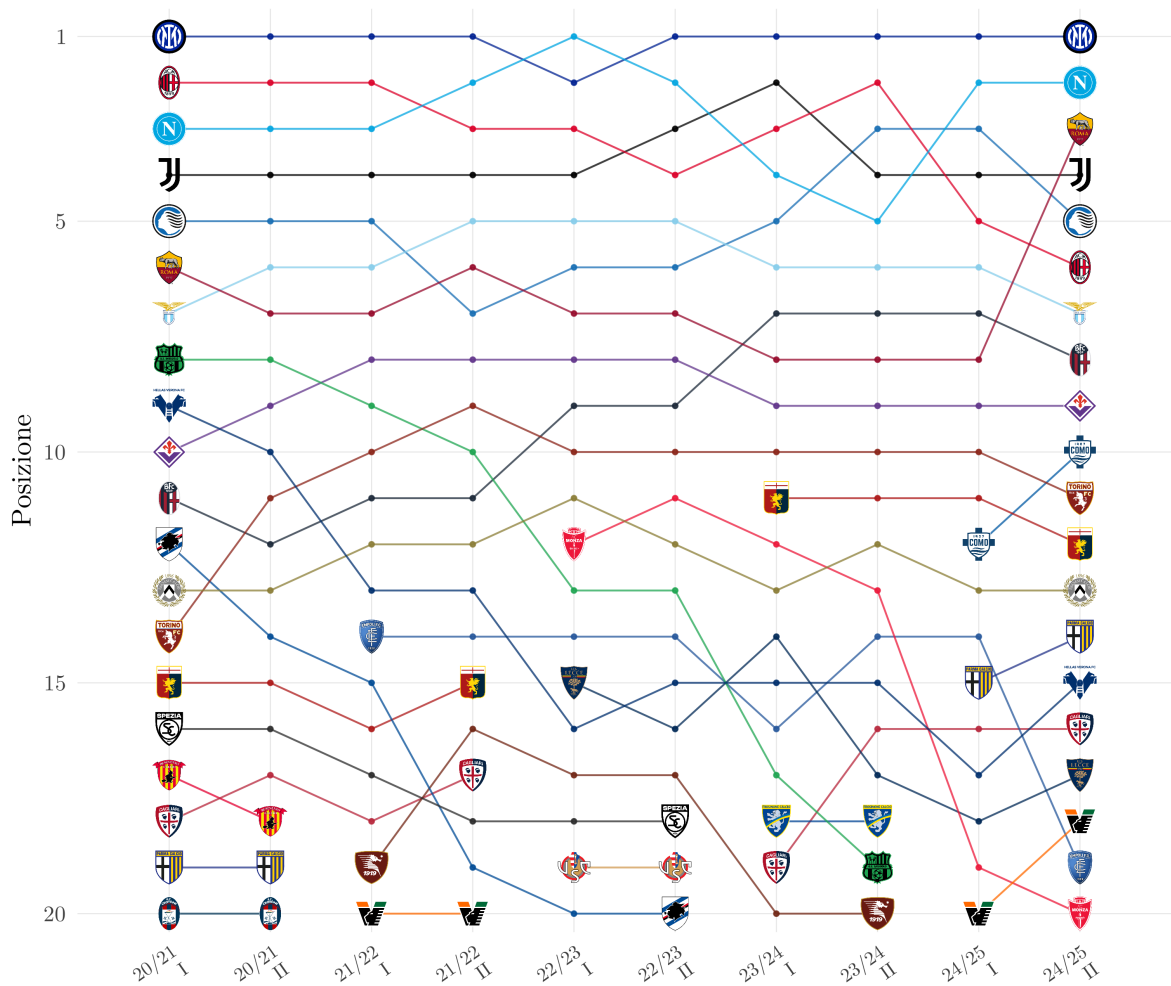


FIGURA 3.2: Evoluzione della posizione nella classifica latente, Serie A, specificazione DOP. Stima sull'intero campione; ciascuna squadra compare soltanto nei periodi in cui ha militato in Serie A.

Le differenze tra gli ordinamenti proposti si concentrano esattamente dove la specificazione adattiva è concepita per intervenire. Un caso che viene posto all'attenzione è quello della Roma nella stagione 2024/25. La squadra è passata da un rendimento da bassa classifica (1.08 punti a partita) nelle prime dodici giornate, ad uno da lotta per i primi posti nel resto del campionato (2.15 punti a partita). Questo cambio di passo è attribuibile agli avvicendamenti della guida tecnica. Nel corso della stagione, dopo gli esoneri di Daniele De Rossi alla quarta giornata e di Ivan Jurić alla dodicesima, il ritorno in panchina di Claudio Ranieri ha permesso alla Roma di conquistare 56 punti in 26 incontri e di chiudere il campionato al quinto posto.

Le tre specificazioni reagiscono a questo cambiamento in modo differente. Nel modello DOP la forza stimata per la Roma registra un incremento di 0.376 nella stagione considerata, salendo dall'ottava alla terza posizione nel *ranking*, mentre nelle due varianti del modello DAOP il medesimo incremento vale rispettivamente 1.11 e 0.97, collocando

la squadra al primo posto della classifica latente al termine della stagione. La ragione risiede nella struttura della varianza di innovazione. Nel modello DAOP la varianza è specifica per squadra e, grazie alla mistura *spike-and-slab*, può variare rapidamente nel tempo. Il meccanismo è esaminato nel paragrafo 3.5.3. Un caso analogo, in senso decrescente, è quello del Monza, la cui forza si riduce di 0.30 nel modello DOP e di circa 0.61 in entrambe le varianti adattive, in corrispondenza della retrocessione maturata con tre giornate di anticipo.

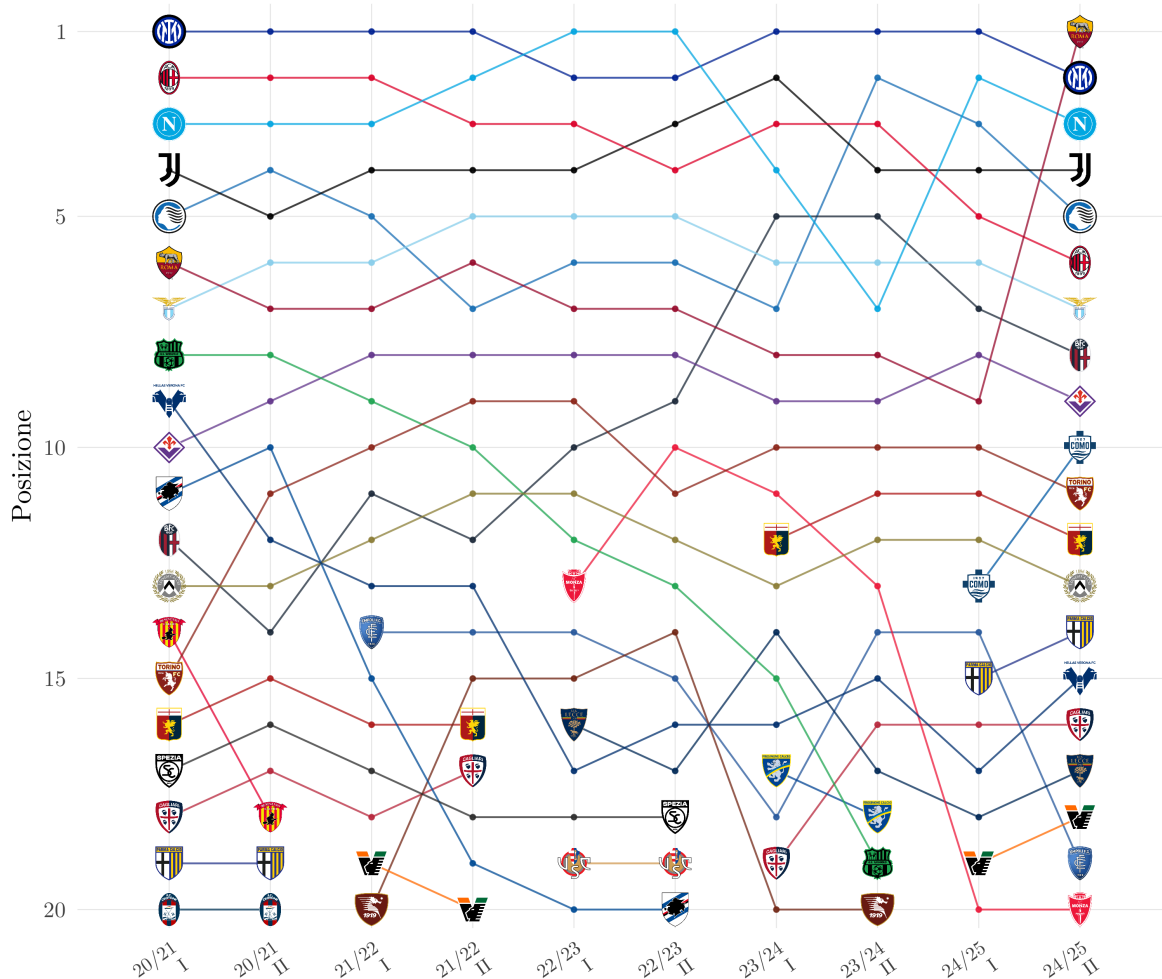


FIGURA 3.3: Evoluzione della posizione nella classifica latente, Serie A, specificazione DAOP con priori $Be(1, 1)$ sul peso della mistura.

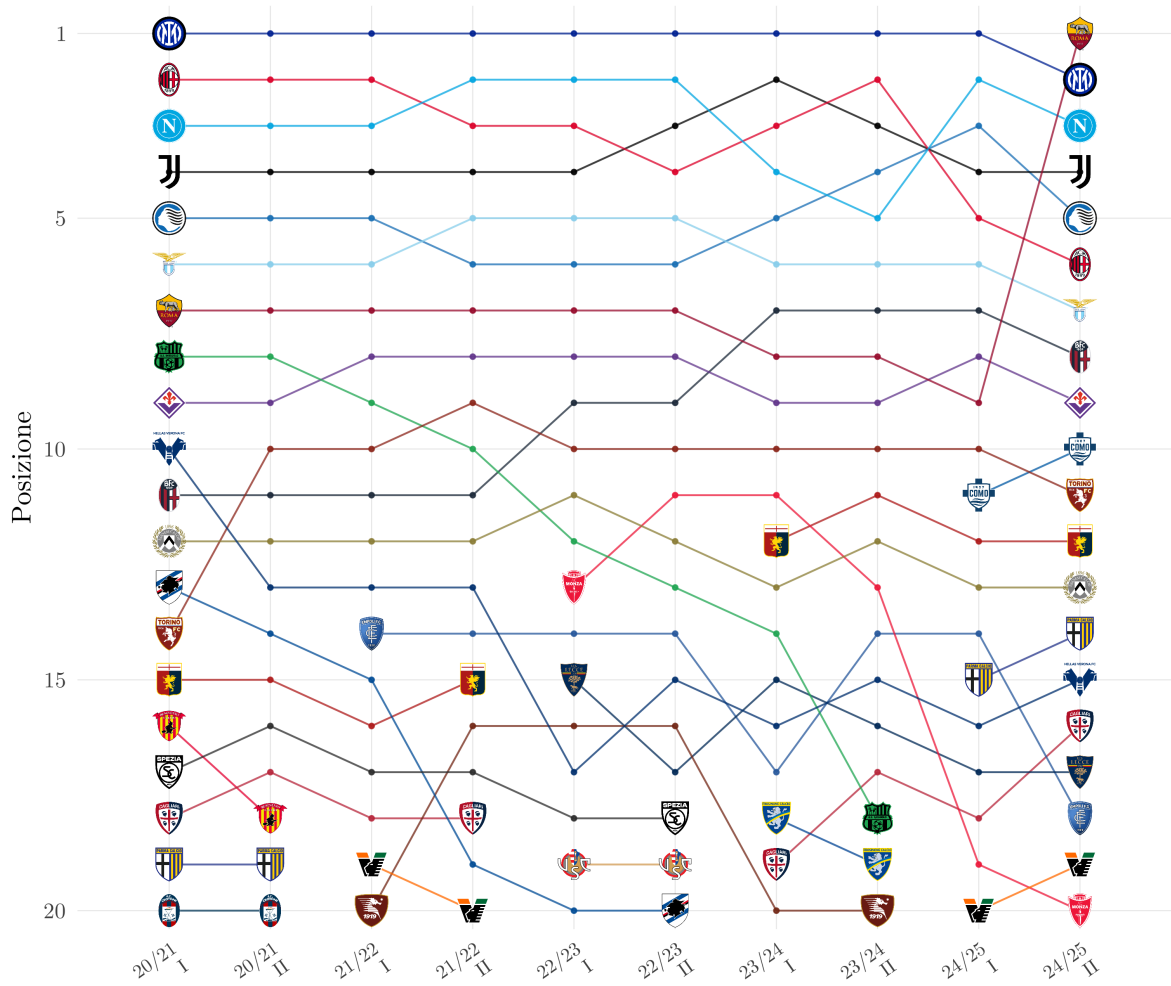


FIGURA 3.4: Evoluzione della posizione nella classifica latente, Serie A, specificazione DAOP con priori $Be(3,1)$ sul peso della mistura.

Il confronto fra le due calibrazioni del modello DAOP mostra invece differenze minime: la priori $Be(3,1)$, più orientata verso la componente *spike*, produce un incremento lievemente inferiore per la Roma, 0.97 contro 1.11, ma l'ordinamento risultante è identico. La maggiore parsimonia si traduce quindi in una moderazione delle variazioni, non in una diversa lettura della gerarchia.

Questi esempi chiariscono che cosa comporti concretamente la scelta della specificazione. Nella grande maggioranza dei casi le due specificazioni coincidono, poiché il rendimento delle squadre varia lentamente e una varianza comune è sufficiente a descriverne l'evoluzione. Tuttavia, quando si verifica una variazione repentina delle prestazioni di una squadra, la specificazione a varianza comune è vincolata dal comportamento delle altre e ne attenua il movimento, mentre quella adattiva lo riconosce e lo accoglie. Le figure relative agli altri tre campionati sono riportate nell'Appendice A.4.

3.5.3 Evoluzione delle forze latenti

In questo paragrafo si analizzano i parametri di forza stimati per le squadre, e si confrontano i meccanismi dinamici delle specificazioni proposte. Le Figure 3.5, 3.7 e 3.9 riportano la distribuzione a posteriori congiunta di $\sigma_{k,t}$ e $\beta_{k,t}$ per tre squadre caratterizzate da dinamiche differenti, in particolare la Lazio, la Roma e il Girona. Per favorire la rappresentazione grafica, sono riportate solamente le ultime tre stagioni, che sono sufficienti a mostrare il differente risultato ottenuto dalle specificazioni dei modelli. Le corrispondenti traiettorie complete delle abilità stimate, in particolare delle mediane per ogni periodo, sono riportate nelle Figure 3.6, 3.8 e 3.10. Nel commento alle figure si fa riferimento alla probabilità a posteriori di appartenenza alla componente diffusa nei modelli adattivi, $P(\xi_{k,t} = 1 \mid y)$, ossia alla frazione di estrazioni in cui la varianza di una squadra è assegnata alla *slab* nel periodo t . Salvo diversa indicazione, i valori numerici di questa quantità riportati nel commento ai tre casi si riferiscono alla calibrazione $\text{Be}(1, 1)$, mentre il confronto fra le due calibrazioni è ripreso al termine del paragrafo.

Il primo caso esaminato è quello della Lazio, riportato nella Figura 3.5, caratterizzato da un rendimento costante nelle cinque stagioni considerate. Per le specificazioni adattive, in tutti i periodi la massa a posteriori si concentra sulla componente *spike*, mentre la componente *slab* riceve un peso quasi nullo e sostanzialmente costante. Coerentemente, la forza latente stimata rimane pressoché immutata attorno a 0.65. In questo caso la specificazione adattiva produce un'escursione complessiva di $\beta_{k,t}$ inferiore a quella della specificazione a varianza comune, pari a 0.140 contro 0.152, e sotto la priori più parsimoniosa essa si riduce ulteriormente a 0.078. Quando i dati non suggeriscono alcun cambiamento nelle prestazioni, la mistura liscia la traiettoria in misura maggiore rispetto alla configurazione più semplice. Le differenze fra le tre specificazioni sono dell'ordine dei centesimi e non risultano percepibili nella Figura 3.6, nella quale le tre curve appaiono sovrapposte.

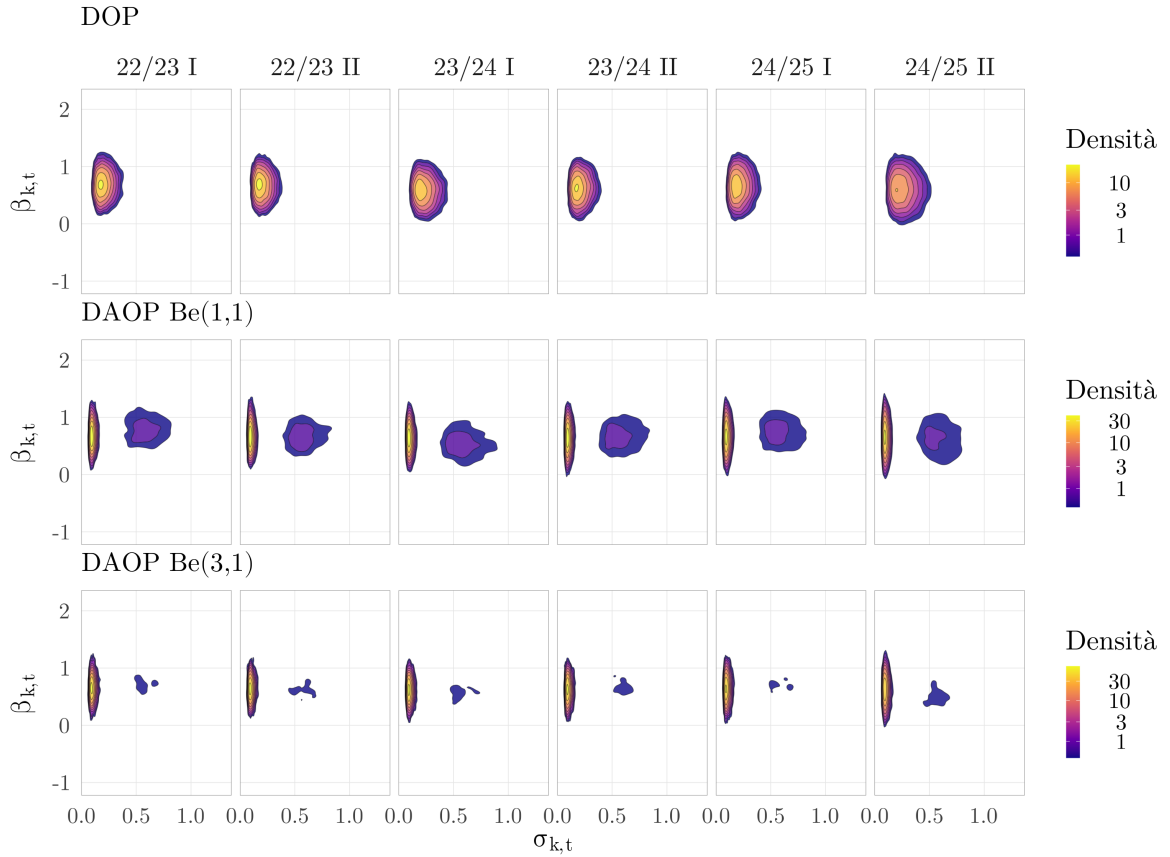


FIGURA 3.5: Distribuzione a posteriori congiunta della deviazione standard di innovazione e della forza latente per la Lazio nelle ultime tre stagioni considerate, per le tre specificazioni dinamiche del modello.

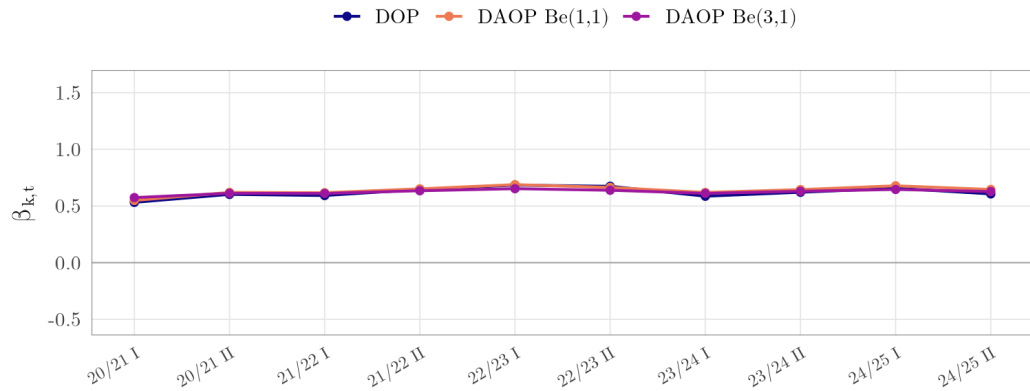


FIGURA 3.6: Mediana a posteriori della forza latente stimata per la Lazio sotto le tre specificazioni, lungo tutte le cinque stagioni in esame.

Il secondo caso in esame è quello della Roma, già discusso nel paragrafo 3.5.2 in relazione all'ordinamento. Come si può osservare dalla Figura 3.7, nei primi periodi si può apprezzare un andamento sostanzialmente costante, con una massa di densità

a posteriori concentrata sulla componente *spike* e una probabilità di appartenenza alla *slab*, sotto la calibrazione $\text{Be}(1, 1)$, compresa fra 0.366 e 0.498. Nell'ultimo periodo, per le medesime cause discusse in precedenza, vi è un cambiamento netto e la probabilità di appartenenza alla *slab* sale a 0.952. La massa di probabilità a posteriori si sposta interamente sulla componente diffusa, in corrispondenza di valori di $\beta_{k,t}$ prossimi a 1.5. La specificazione a varianza comune del modello DOP, vincolata dal comportamento delle altre squadre, produce nel medesimo periodo uno spostamento verso l'alto della massa di probabilità a posteriori più attenuato rispetto alle varianti adattive.

Dalla Figura 3.8 si può osservare che nel penultimo periodo le due specificazioni adattive scendono a 0.36 e 0.42, mentre nel modello DOP sale a 0.58. Le prime seguono dunque anche il rendimento negativo della prima parte della stagione, che la varianza comune attenua, prima di registrarne il ribaltamento nel periodo conclusivo.

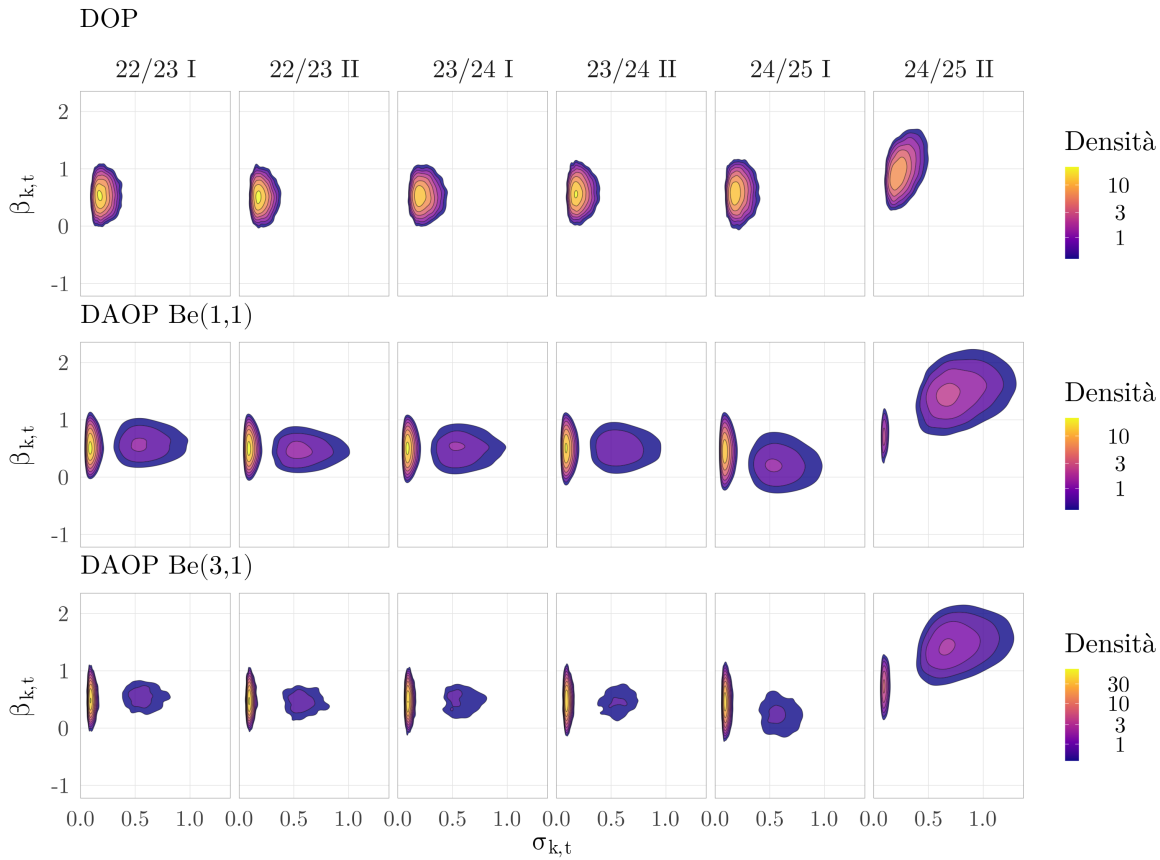


FIGURA 3.7: Distribuzione a posteriori congiunta della deviazione standard di innovazione e della forza latente per la Roma nelle ultime tre stagioni considerate, per le tre specificazioni dinamiche del modello.

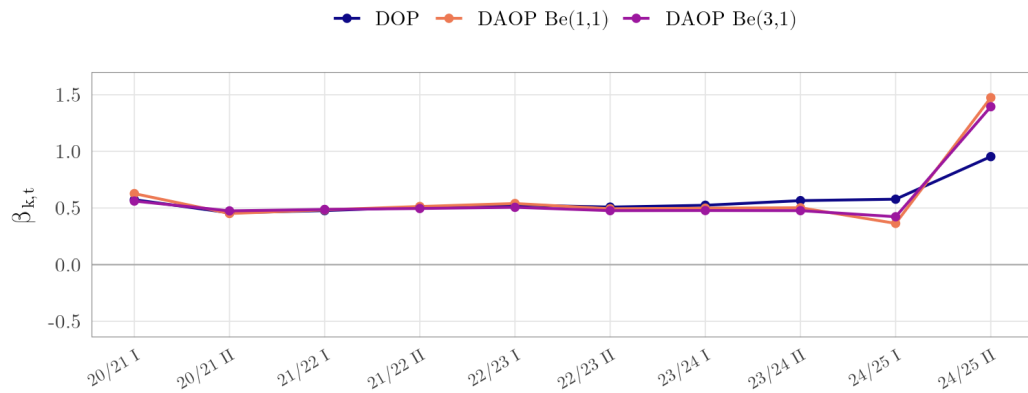


FIGURA 3.8: Mediana a posteriori della forza latente stimata per la Roma sotto le tre specificazioni, lungo tutte le cinque stagioni in esame.

Il terzo caso, riportato nella Figura 3.9, è quello del Girona, e presenta un comportamento diverso dai precedenti. La squadra, promossa nella massima divisione spagnola a partire dalla stagione 2022/23, nel 2023/24 ha avuto un *exploit* concludendo il campionato al terzo posto con 81 punti, qualificandosi per la prima volta nella propria storia alla UEFA Champions League. Nella stagione successiva ha registrato tuttavia un tracollo nelle prestazioni, riuscendo a chiudere il campionato soltanto al sedicesimo posto. Come si può apprezzare dalla Figura 3.9, in questo caso la componente *slab* domina in ogni periodo, con una probabilità a posteriori mai inferiore a 0.718 e pari in media a 0.838. Il modello non registra dunque un episodio isolato, ma classifica il Girona come squadra permanentemente instabile, e ne asseconda i cambiamenti di prestazioni in entrambe le direzioni. La specificazione a varianza comune segue lo stesso percorso in forma attenuata, con un'escursione della forza latente stimata pari a 0.695.

La Figura 3.10 mostra le traiettorie stimate dai modelli, raffigurando un modo via via più netto di rappresentare la dinamica della forza latente. La specificazione che lascia di più la traiettoria in questo caso è il modello DOP, mentre la più reattiva ai cambiamenti è il modello DAOP. In particolare la versione più parsimoniosa sembra avere un comportamento intermedio rispetto agli altri due casi. Nel modello DAOP con priori $\text{Be}(1,1)$ la forza latente stimata sale da 0.05 a 1.16 e scende fino a -0.42 , con un'escursione complessiva di 1.576, la più ampia osservata in tutti e quattro i campionati considerati.

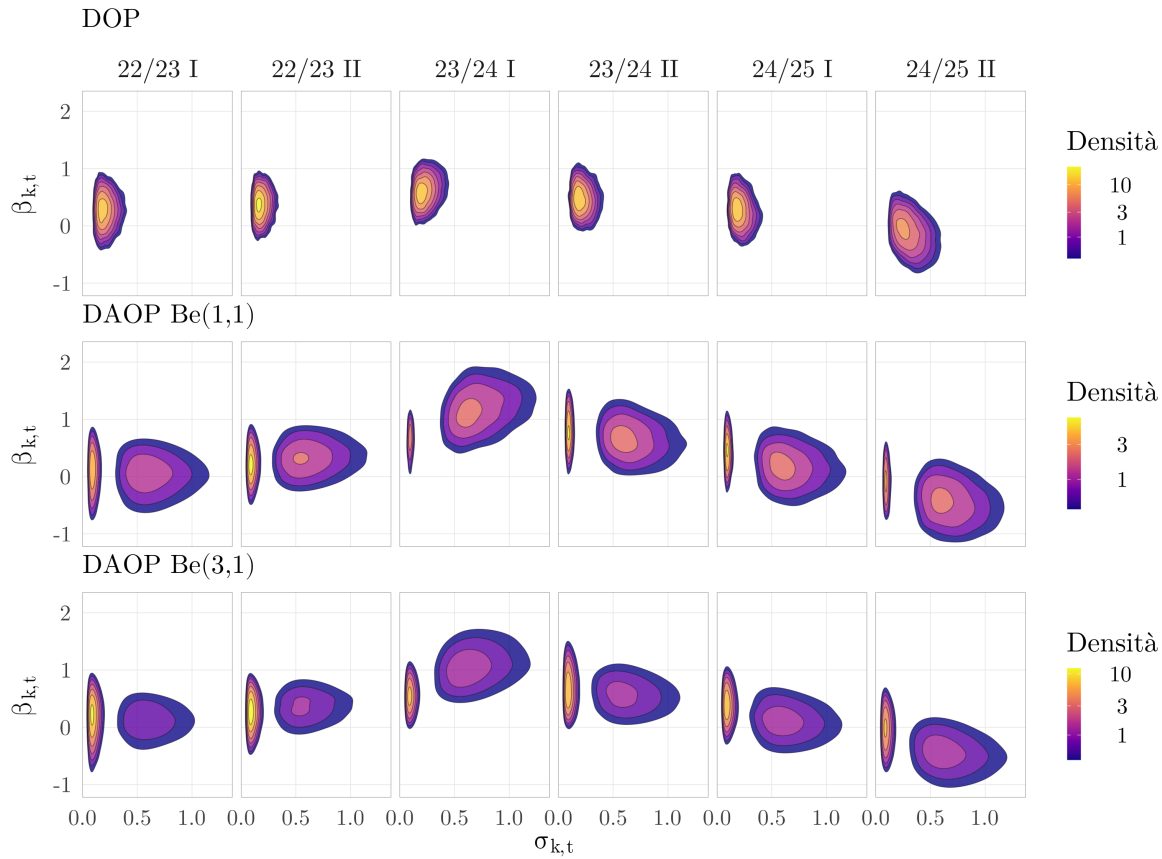


FIGURA 3.9: Distribuzione a posteriori congiunta della deviazione standard di innovazione e della forza latente per il Girona nelle stagioni in cui ha preso parte alla Liga spagnola, per le tre specificazioni dinamiche del modello.

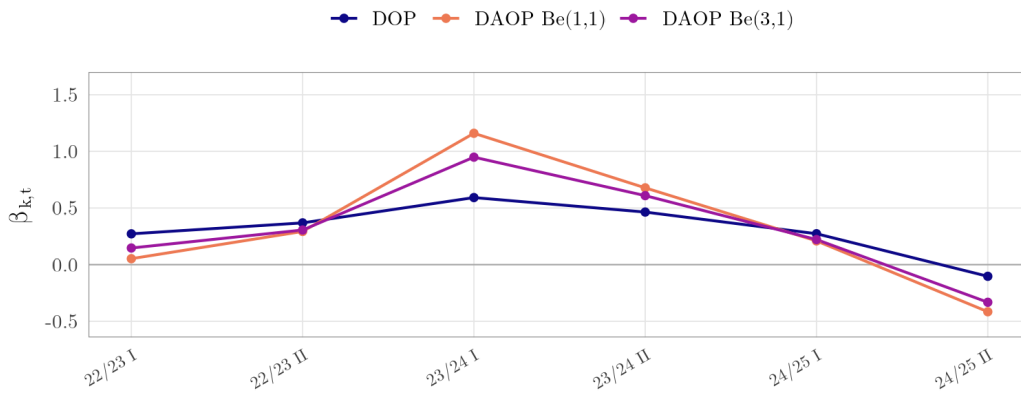


FIGURA 3.10: Mediana a posteriori della forza latente stimata per il Girona sotto le tre specificazioni, nelle stagioni in cui ha preso parte alla Liga spagnola.

Il confronto fra le due calibrazioni del modello DAOP conferma quanto osservato in sede di analisi degli ordinamenti. La priori $Be(3, 1)$, più orientata verso la componente concentrata, riduce sistematicamente la probabilità di appartenenza alla *slab* rispetto

alla versione meno conservativa. La sua media in tutti i periodi passa da 0.216 a 0.087 per la Lazio, da 0.460 a 0.251 per la Roma e da 0.838 a 0.614 per il Girona. Le escursioni delle forze latenti si contraggono di conseguenza, ma la graduatoria dei tre casi resta invariata e le variazioni principali continuano a essere riconosciute. Il segnale contenuto nei dati risulta quindi sufficientemente forte da manifestarsi anche sotto la distribuzione a priori che favorisce l'ipotesi di assenza di cambiamento.

Il peso della mistura p_k è l'unico parametro della specificazione adattiva che non varia nel tempo, e caratterizza l'intera dinamica di ogni compagine. Infatti, valori elevati del peso della componente diffusa, $1 - p_k$, corrispondono a squadre per le quali il modello ammette con frequenza innovazioni di grande ampiezza. La Figura 3.11 riporta la distribuzione a posteriori di $1 - p_k$ per le squadre della Serie A presenti in almeno cinque periodi, sotto entrambe le calibrazioni. Considerando le mediane a posteriori delle distribuzioni di p_k si può ottenere un ordinamento rispetto alla instabilità nelle prestazioni delle squadre. In Serie A, per entrambe le specificazioni adattive, i valori più elevati di $1 - p_k$ spettano a Monza e Salernitana, il più basso alla Lazio che è considerata dunque la squadra più stabile. Il valore mediano massimo fra tutti e quattro i campionati è quello del Girona, pari a 0.786. Le figure relative agli altri tre campionati sono riportate nell'Appendice A.5.

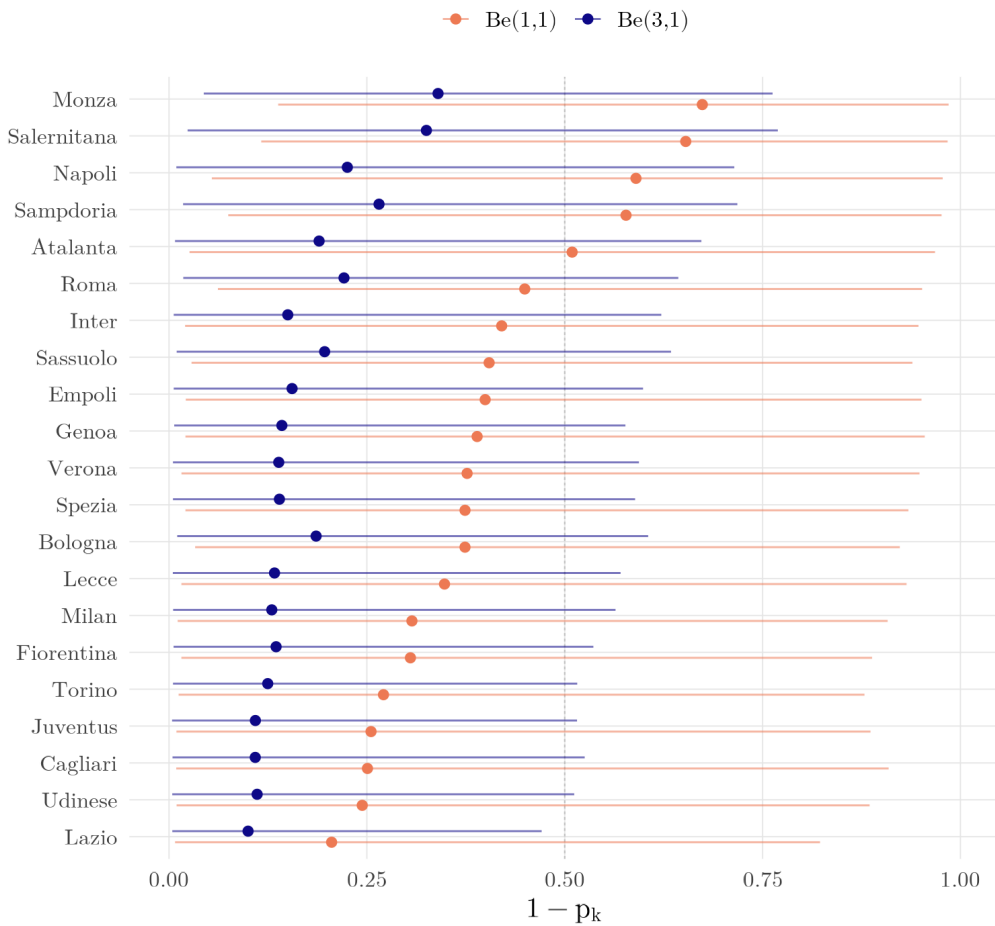


FIGURA 3.11: Distribuzione a posteriori del peso della componente *slab*, $1 - p_k$, per le squadre della Serie A presenti in almeno cinque periodi, sotto le due calibrazioni della priori su p_k . Il punto rappresenta la mediana a posteriori, il segmento l'intervallo di credibilità al 95%. Le squadre sono ordinate secondo la mediana ottenuta sotto la calibrazione $\text{Be}(1, 1)$.

3.6 Analisi predittiva

3.6.1 Confronto con i modelli *goal-based*

Le Tabelle 3.2, 3.3 e 3.4 riportano, per ciascuno dei tre scenari descritti nel paragrafo 3.2, i valori assunti dalle quattro metriche introdotte nel paragrafo 3.4 dalle specificazioni proposte in questa tesi, dai cinque modelli *goal-based* presentati in Macrì-Demartino et al. (2026) e dal riferimento di mercato. Per ogni combinazione di campionato e scenario, il grassetto individua il modello migliore rispetto a ciascuna metrica, escludendo il mercato.

TABELLA 3.2: Metriche di previsione per la seconda parte di stagione 2023/24. RPS, Brier score, ACP e pseudo- R^2 per lega e modello. Il grassetto individua il modello migliore rispetto a ciascuna metrica, escludendo il mercato (indicato in corsivo).

Lega	Modello	n	RPS	Brier	ACP	pseudo- R^2
Bundesliga	SOP	163	0.2105	0.6244	0.3984	0.3521
	DOP	163	0.1983	0.6031	0.4261	0.3651
	DAOP Be(1,1)	163	0.1953	0.5986	0.4400	0.3682
	DAOP Be(3,1)	163	0.1950	0.5964	0.4351	0.3691
	Poisson	163	0.1970	0.5991	0.4153	0.3676
	Dixon-Coles	163	0.1976	0.6000	0.4150	0.3674
	Binomiale negativa	163	0.1975	0.5998	0.4087	0.3671
	Skellam	163	0.1970	0.5971	0.4016	0.3687
	ZI Skellam	163	0.1983	0.6007	0.4004	0.3666
	<i>Mercato</i>	<i>163</i>	<i>0.1887</i>	<i>0.5796</i>	<i>0.4245</i>	<i>0.3770</i>
LaLiga	SOP	200	0.1984	0.5839	0.4166	0.3760
	DOP	200	0.2022	0.5919	0.4334	0.3710
	DAOP Be(1,1)	200	0.2053	0.5995	0.4400	0.3657
	DAOP Be(3,1)	200	0.2027	0.5934	0.4358	0.3699
	Poisson	200	0.1975	0.5809	0.4094	0.3774
	Dixon-Coles	200	0.1980	0.5818	0.4082	0.3767
	Binomiale negativa	200	0.1985	0.5830	0.4047	0.3761
	Skellam	200	0.2039	0.5962	0.3846	0.3691
	ZI Skellam	200	0.2034	0.5950	0.3857	0.3697
	<i>Mercato</i>	<i>200</i>	<i>0.1902</i>	<i>0.5670</i>	<i>0.4168</i>	<i>0.3850</i>
Premier	SOP	182	0.1722	0.5274	0.4553	0.4057
	DOP	182	0.1749	0.5341	0.4608	0.4034
	DAOP Be(1,1)	182	0.1783	0.5423	0.4615	0.3986
	DAOP Be(3,1)	182	0.1754	0.5354	0.4618	0.4026
	Poisson	182	0.1803	0.5431	0.4376	0.3981
	Dixon-Coles	182	0.1800	0.5424	0.4380	0.3984
	Binomiale negativa	182	0.1808	0.5442	0.4323	0.3969
	Skellam	182	0.1891	0.5618	0.4103	0.3856
	ZI Skellam	182	0.1880	0.5594	0.4124	0.3869
	<i>Mercato</i>	<i>182</i>	<i>0.1677</i>	<i>0.5140</i>	<i>0.4629</i>	<i>0.4137</i>
Serie A	SOP	210	0.2002	0.6229	0.3975	0.3577
	DOP	210	0.1995	0.6254	0.4058	0.3578
	DAOP Be(1,1)	210	0.1986	0.6259	0.4120	0.3567
	DAOP Be(3,1)	210	0.1989	0.6244	0.4073	0.3581
	Poisson	210	0.1937	0.6065	0.4003	0.3662
	Dixon-Coles	210	0.1936	0.6064	0.3999	0.3662
	Binomiale negativa	210	0.1941	0.6070	0.3964	0.3657
	Skellam	210	0.1937	0.6040	0.3880	0.3653
	ZI Skellam	210	0.1940	0.6049	0.3874	0.3650
	<i>Mercato</i>	<i>210</i>	<i>0.1862</i>	<i>0.5923</i>	<i>0.4064</i>	<i>0.3733</i>

TABELLA 3.3: Metriche di previsione per la seconda parte di stagione 2024/25. RPS, Brier *score*, ACP e pseudo- R^2 per lega e modello. Il grassetto individua il modello migliore rispetto a ciascuna metrica, escludendo il mercato (indicato in corsivo).

Lega	Modello	n	RPS	Brier	ACP	pseudo- R^2
Bundesliga	SOP	171	0.2205	0.6392	0.3944	0.3444
	DOP	171	0.2172	0.6353	0.4052	0.3477
	DAOP Be(1,1)	171	0.2178	0.6380	0.4094	0.3462
	DAOP Be(3,1)	171	0.2182	0.6380	0.4060	0.3463
	Poisson	171	0.2133	0.6267	0.3905	0.3530
	Dixon-Coles	171	0.2124	0.6251	0.3915	0.3538
	Binomiale negativa	171	0.2116	0.6232	0.3868	0.3548
	Skellam	171	0.2143	0.6279	0.3753	0.3517
	ZI Skellam	171	0.2148	0.6294	0.3748	0.3509
	<i>Mercato</i>	<i>171</i>	<i>0.2118</i>	<i>0.6195</i>	<i>0.3979</i>	<i>0.3570</i>
LaLiga	SOP	199	0.2089	0.6005	0.4093	0.3652
	DOP	199	0.2054	0.5919	0.4249	0.3694
	DAOP Be(1,1)	199	0.2051	0.5912	0.4270	0.3699
	DAOP Be(3,1)	199	0.2056	0.5923	0.4233	0.3692
	Poisson	199	0.2005	0.5836	0.4112	0.3743
	Dixon-Coles	199	0.2006	0.5841	0.4112	0.3739
	Binomiale negativa	199	0.2010	0.5850	0.4063	0.3733
	Skellam	199	0.2038	0.5907	0.3916	0.3707
	ZI Skellam	199	0.2034	0.5895	0.3923	0.3713
	<i>Mercato</i>	<i>199</i>	<i>0.1880</i>	<i>0.5549</i>	<i>0.4316</i>	<i>0.3906</i>
Premier	SOP	181	0.2177	0.6003	0.4070	0.3680
	DOP	181	0.2043	0.5734	0.4341	0.3828
	DAOP Be(1,1)	181	0.2053	0.5752	0.4401	0.3818
	DAOP Be(3,1)	181	0.2057	0.5759	0.4341	0.3811
	Poisson	181	0.2079	0.5792	0.4130	0.3790
	Dixon-Coles	181	0.2086	0.5806	0.4129	0.3783
	Binomiale negativa	181	0.2083	0.5797	0.4093	0.3787
	Skellam	181	0.2148	0.5958	0.3905	0.3693
	ZI Skellam	181	0.2149	0.5968	0.3898	0.3688
	<i>Mercato</i>	<i>181</i>	<i>0.2003</i>	<i>0.5653</i>	<i>0.4296</i>	<i>0.3861</i>
Serie A	SOP	212	0.1992	0.6083	0.4109	0.3641
	DOP	212	0.1986	0.6068	0.4107	0.3643
	DAOP Be(1,1)	212	0.1988	0.6084	0.4141	0.3635
	DAOP Be(3,1)	212	0.1978	0.6054	0.4123	0.3652
	Poisson	212	0.1969	0.6023	0.4005	0.3657
	Dixon-Coles	212	0.1971	0.6026	0.4010	0.3656
	Binomiale negativa	212	0.1973	0.6030	0.3970	0.3654
	Skellam	212	0.2034	0.6159	0.3805	0.3583
	ZI Skellam	212	0.2031	0.6151	0.3812	0.3586
	<i>Mercato</i>	<i>212</i>	<i>0.1816</i>	<i>0.5685</i>	<i>0.4205</i>	<i>0.3849</i>

TABELLA 3.4: Metriche di previsione per l'ultima giornata della stagione 2024/25. RPS, Brier *score*, ACP e pseudo- R^2 per lega e modello. Il grassetto individua il modello migliore rispetto a ciascuna metrica, escludendo il mercato (indicato in corsivo).

Lega	Modello	n	RPS	Brier	ACP	pseudo- R^2
Bundesliga	SOP	9	0.2582	0.6311	0.3908	0.3515
	DOP	9	0.2098	0.5421	0.4381	0.4046
	DAOP Be(1,1)	9	0.2121	0.5427	0.4372	0.4037
	DAOP Be(3,1)	9	0.2106	0.5443	0.4363	0.4033
	Poisson	9	0.2262	0.5705	0.4069	0.3802
	Dixon-Coles	9	0.2237	0.5677	0.4097	0.3826
	Binomiale negativa	9	0.2245	0.5679	0.4066	0.3827
	Skellam	9	0.2319	0.5788	0.3944	0.3770
	ZI Skellam	9	0.2320	0.5789	0.3943	0.3773
	<i>Mercato</i>	<i>9</i>	<i>0.2609</i>	<i>0.6291</i>	<i>0.3947</i>	<i>0.3559</i>
LaLiga	SOP	10	0.1588	0.5251	0.4318	0.4105
	DOP	10	0.1229	0.4424	0.4956	0.4598
	DAOP Be(1,1)	10	0.1141	0.4220	0.5155	0.4728
	DAOP Be(3,1)	10	0.1205	0.4358	0.4996	0.4632
	Poisson	10	0.1504	0.4927	0.4484	0.4259
	Dixon-Coles	10	0.1480	0.4872	0.4509	0.4288
	Binomiale negativa	10	0.1589	0.5119	0.4327	0.4139
	Skellam	10	0.1734	0.5264	0.4153	0.4070
	ZI Skellam	10	0.1704	0.5198	0.4191	0.4108
	<i>Mercato</i>	<i>10</i>	<i>0.1223</i>	<i>0.4430</i>	<i>0.4964</i>	<i>0.4596</i>
Premier	SOP	10	0.1990	0.5624	0.4314	0.3891
	DOP	10	0.1969	0.5533	0.4480	0.3954
	DAOP Be(1,1)	10	0.1976	0.5525	0.4549	0.3951
	DAOP Be(3,1)	10	0.1985	0.5556	0.4489	0.3944
	Poisson	10	0.2010	0.5639	0.4243	0.3869
	Dixon-Coles	10	0.1984	0.5584	0.4284	0.3903
	Binomiale negativa	10	0.1994	0.5602	0.4230	0.3885
	Skellam	10	0.1977	0.5504	0.4188	0.3957
	ZI Skellam	10	0.1988	0.5522	0.4172	0.3945
	<i>Mercato</i>	<i>10</i>	<i>0.2232</i>	<i>0.6086</i>	<i>0.4193</i>	<i>0.3546</i>
Serie A	SOP	10	0.2755	0.6096	0.4344	0.3346
	DOP	10	0.2809	0.6149	0.4386	0.3305
	DAOP Be(1,1)	10	0.2747	0.5968	0.4601	0.3411
	DAOP Be(3,1)	10	0.2759	0.6015	0.4559	0.3370
	Poisson	10	0.2627	0.5900	0.4186	0.3593
	Dixon-Coles	10	0.2643	0.5928	0.4184	0.3572
	Binomiale negativa	10	0.2639	0.5933	0.4133	0.3590
	Skellam	10	0.2537	0.5935	0.3937	0.3651
	ZI Skellam	10	0.2512	0.5853	0.3982	0.3699
	<i>Mercato</i>	<i>10</i>	<i>0.2536</i>	<i>0.5599</i>	<i>0.4534</i>	<i>0.3729</i>

Dal confronto fra i due approcci non emerge un vantaggio sistematico di una delle due classi di modelli. Limitando l'attenzione ai primi due scenari, nelle otto combinazioni tra campionato e periodo il valore di RPS più basso è ottenuto tre volte da una specificazione *result-based* e cinque volte da un modello *goal-based*. I modelli che considerano unicamente l'esito qualitativo dell'incontro risultano pertanto competitivi con specificazioni che dispongono dell'informazione più ricca contenuta nel punteggio.

Il commento che segue si concentra sui primi due scenari, la cui numerosità consente un confronto più informativo fra i modelli, mentre lo scenario relativo all'ultima giornata viene ripreso in chiusura del paragrafo. Le metriche sono calcolate come medie sui relativi insiemi di incontri e, come è consueto nella letteratura sulla previsione dei risultati calcistici, sono presentate in forma puntuale.

Si considera anzitutto il confronto interno alle specificazioni proposte. La specificazione costante presenta generalmente performance predittive inferiori rispetto alle varianti dinamiche. Nello specifico, in Bundesliga e Serie A è la peggiore delle quattro in entrambi i periodi, con un RPS di 0.2105 e 0.2205 contro lo 0.1950 e lo 0.2172 della migliore fra le altre, mentre per Premier League e Liga ottiene il risultato migliore nella seconda parte della stagione 2023/24, e il peggiore nella seconda parte della stagione successiva. In Serie A risulta in entrambi i casi il peggiore. Fra il modello DOP e le due calibrazioni del modello DAOP le differenze sono invece dell'ordine dei millesimi e nessuna delle tre prevale con regolarità nei due scenari in esame.

Il confronto con i modelli *goal-based* restituisce prestazioni comparabili. La Premier League è l'unico campionato nel quale le specificazioni dinamiche proposte risultano migliori di tutti i modelli *goal-based*, in entrambi gli scenari. Negli altri tre campionati il primato spetta più frequentemente ai modelli *goal-based*, con distacchi di pochi millesimi sul RPS e più ampi sul Brier *score*. Le specificazioni proposte utilizzano esclusivamente l'esito dell'incontro, mentre i modelli di confronto dispongono del punteggio e modellano separatamente le componenti di attacco e difesa. L'assenza di un divario sistematico, e i risultati talvolta preferibili per i modelli ordinali a fronte di una disponibilità informativa più limitata costituisce pertanto l'elemento di maggiore interesse del confronto.

La media delle probabilità corrette restituisce un quadro più regolare, e lo fa rispetto a tutti i confronti considerati. In ciascuno degli otto casi le tre specificazioni dinamiche proposte registrano un ACP superiore a quello di ogni modello *goal-based*. La specificazione statica si colloca quasi sempre al di sotto delle altre tre, e fra queste ultime una delle due calibrazioni del modello DAOP risulta sempre preferibile al modello DOP. In questo caso, nel confronto con il riferimento di mercato, la migliore specificazione proposta lo supera in cinque casi su otto. Lo pseudo- R^2 riproduce quasi esattamente

l'ordinamento del RPS e del Brier *score*.

Il riferimento di mercato resta il previsore più accurato in termini di RPS, ottenendo il valore più basso in sette delle otto combinazioni considerate. L'unica eccezione si registra in Bundesliga nella seconda parte della stagione 2024/25, dove è superato dal modello binomiale negativo. Il quadro muta se si include lo scenario dell'ultima giornata, nel quale il mercato viene battuto in tutti e quattro i campionati, in tre dei quali da una delle specificazioni proposte nella tesi. Tale scenario comprende però soltanto nove o dieci incontri per campionato, cosicché le metriche risultano assai più sensibili alla variabilità campionaria, e presenta caratteristiche peculiari legate sia alla diversa rilevanza delle partite ai fini della classifica sia alle motivazioni differenti delle squadre, rendendo questo scenario molto imprevedibile.

Si indaga ora l'allocazione delle probabilità degli esiti previsti indotta dalla struttura del modello. Con soglie fissate, le tre probabilità previste dipendono dalla partita attraverso il predittore lineare η_i e la varianza dell'errore σ_ε , in questo modo le previsioni si concentrano attorno ad una curva nel simplex di probabilità con una dispersione data da σ_ε , riportato nella Figura 3.12. La probabilità del pareggio stimata dai modelli presenta un valore massimo di 0.31 nei quattro campionati, pertanto il modello non assegna mai al pareggio la probabilità più alta. Tuttavia si può notare che anche le probabilità implicite nelle quote proposte dal mercato si dispongono in un modo analogo e indicano il pareggio come esito più probabile in appena sei incontri su oltre settemila considerati.

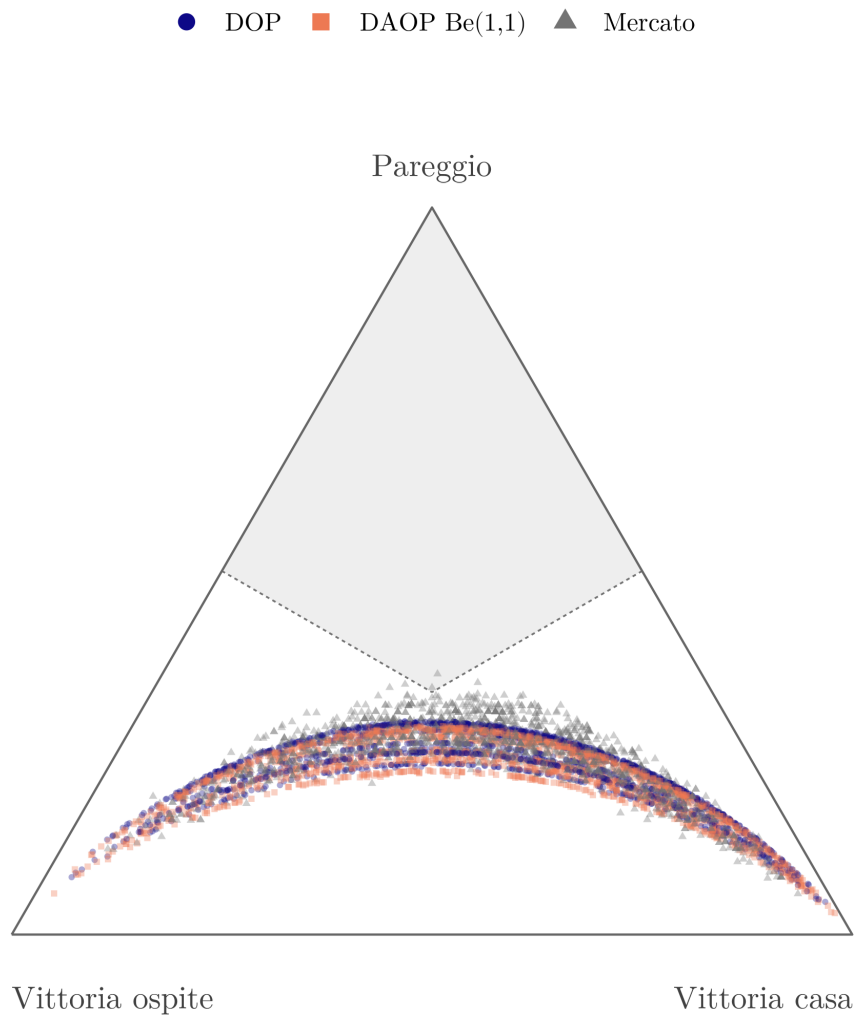


FIGURA 3.12: Previsioni dei modelli DOP e DAOP con priori $Be(1, 1)$ e probabilità implicite nelle quote di mercato, rappresentate nel semplice dei tre esiti sui due scenari a metà stagione. Le previsioni che ricadono nella zona ombreggiata indicano il pareggio come l'esito più probabile.

La Figura 3.13 rappresenta le distribuzioni del RPS distinguendo per esito osservato, sulle 1 518 partite previste nei primi due scenari nei quattro campionati. In questo modo è possibile comprendere in che modo tale limite strutturale si traduce in perdita di capacità predittiva. Il divario del modello DOP rispetto al mercato vale 0.0038 sulle vittorie interne e 0.0114 su quelle esterne, ma sale a 0.0216 sui pareggi, i quali, pur costituendo il 27% degli incontri, causano il 52% del divario complessivo. All'origine vi è la sottostima sistematica del pareggio, al quale il modello DOP assegna in media una probabilità di 0.227 contro lo 0.245 delle quote, a fronte di una frequenza osservata di

0.266. Il modello DAOP produce previsioni più nette e riduce ulteriormente la probabilità attribuita al pareggio, che scende a 0.215. Ne consegue un punteggio migliore di quello del modello DOP su entrambi gli esiti di vittoria, 0.2589 contro 0.2631 in trasferta e 0.1575 contro 0.1582 in casa, e peggiore sui pareggi, 0.1945 contro 0.1863. I due effetti si compensano quasi esattamente, cosicché il RPS complessivo passa da 0.2000 a 0.2005, mentre la quota del divario dal mercato imputabile ai pareggi sale dal 52% al 69%. Questo modo più netto di produrre le previsioni giustifica l'ACP più elevato rilevato in precedenza, dal momento che le vittorie costituiscono la maggior parte degli esiti osservati.

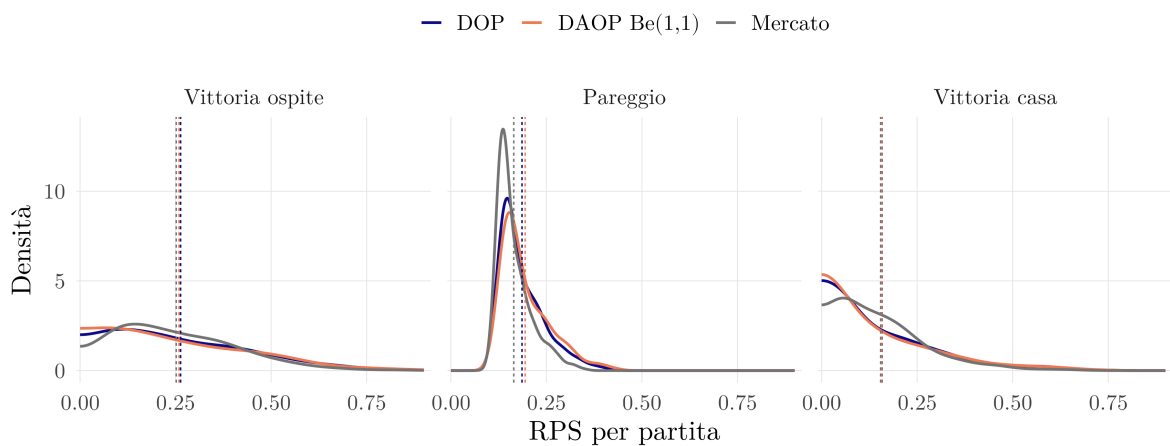


FIGURA 3.13: Distribuzione del RPS per partita, separata per esito osservato, per i modelli DOP e DAOP con priori $\text{Be}(1,1)$, e per il riferimento di mercato. Le linee tratteggiate indicano le medie.

3.6.2 Validazione a finestra crescente

Le Tabelle 3.5–3.7 riportano i risultati dello schema a finestra crescente descritto nel paragrafo 3.2 per $H \in \{1, 5, 10\}$, per le specificazioni dinamiche del modello ordinale. Il confronto con il riferimento di mercato è possibile solamente per $H = 1$.

TABELLA 3.5: Metriche predittive per l'orizzonte $H = 1$. RPS, Brier score, ACP e pseudo- R^2 medi per lega e specificazione. Il grassetto indica il miglior modello per la relativa metrica, escludendo il riferimento di mercato (indicato in corsivo).

Lega	Modello	n	RPS	Brier	ACP	pseudo- R^2
Serie A	DOP	172	0.1923	0.5809	0.4280	0.3778
	DAOP Be(1,1)	172	0.1901	0.5777	0.4411	0.3800
	DAOP Be(3,1)	172	0.1895	0.5752	0.4366	0.3811
	<i>Mercato</i>	<i>172</i>	<i>0.1822</i>	<i>0.5570</i>	<i>0.4270</i>	<i>0.3909</i>
Bundesliga	DOP	126	0.2267	0.6714	0.3810	0.3307
	DAOP Be(1,1)	126	0.2314	0.6836	0.3798	0.3233
	DAOP Be(3,1)	126	0.2284	0.6755	0.3796	0.3286
	<i>Mercato</i>	<i>126</i>	<i>0.2205</i>	<i>0.6503</i>	<i>0.3809</i>	<i>0.3416</i>
Premier	DOP	133	0.2031	0.5850	0.4295	0.3761
	DAOP Be(1,1)	133	0.2057	0.5912	0.4348	0.3721
	DAOP Be(3,1)	133	0.2055	0.5900	0.4296	0.3730
	<i>Mercato</i>	<i>133</i>	<i>0.1953</i>	<i>0.5652</i>	<i>0.4285</i>	<i>0.3851</i>
LaLiga	DOP	160	0.1877	0.5524	0.4368	0.3924
	DAOP Be(1,1)	160	0.1868	0.5505	0.4434	0.3935
	DAOP Be(3,1)	160	0.1874	0.5511	0.4380	0.3929
	<i>Mercato</i>	<i>160</i>	<i>0.1815</i>	<i>0.5380</i>	<i>0.4393</i>	<i>0.4005</i>

TABELLA 3.6: Metriche predittive per l'orizzonte $H = 5$. RPS, Brier score, ACP e pseudo- R^2 medi per lega e specificazione. Il grassetto indica il miglior modello per la relativa metrica.

Lega	Modello	n	RPS	Brier	ACP	pseudo- R^2
Serie A	DOP	665	0.1900	0.5842	0.4258	0.3773
	DAOP Be(1,1)	665	0.1869	0.5793	0.4394	0.3803
	DAOP Be(3,1)	665	0.1872	0.5787	0.4335	0.3805
Bundesliga	DOP	450	0.2402	0.6985	0.3721	0.3158
	DAOP Be(1,1)	450	0.2505	0.7221	0.3677	0.3030
	DAOP Be(3,1)	450	0.2435	0.7056	0.3698	0.3124
Premier	DOP	512	0.2008	0.5896	0.4283	0.3742
	DAOP Be(1,1)	512	0.2037	0.5977	0.4327	0.3694
	DAOP Be(3,1)	512	0.2028	0.5943	0.4288	0.3715
LaLiga	DOP	601	0.1933	0.5659	0.4309	0.3851
	DAOP Be(1,1)	601	0.1932	0.5662	0.4358	0.3850
	DAOP Be(3,1)	601	0.1934	0.5656	0.4313	0.3850

TABELLA 3.7: Metriche predittive per l'orizzonte $H = 10$. RPS, Brier score, ACP e pseudo- R^2 medi per lega e specificazione. Il grassetto indica il miglior modello per la relativa metrica.

Lega	Modello	n	RPS	Brier	ACP	pseudo- R^2
Serie A	DOP	892	0.1866	0.5832	0.4242	0.3778
	DAOP Be(1,1)	892	0.1828	0.5775	0.4377	0.3806
	DAOP Be(3,1)	892	0.1842	0.5784	0.4306	0.3803
Bundesliga	DOP	450	0.2422	0.7037	0.3763	0.3125
	DAOP Be(1,1)	450	0.2553	0.7342	0.3726	0.2957
	DAOP Be(3,1)	450	0.2473	0.7153	0.3737	0.3068
Premier	DOP	593	0.2041	0.5963	0.4235	0.3707
	DAOP Be(1,1)	593	0.2060	0.6028	0.4285	0.3667
	DAOP Be(3,1)	593	0.2057	0.6004	0.4243	0.3685
LaLiga	DOP	770	0.2000	0.5792	0.4269	0.3774
	DAOP Be(1,1)	770	0.1996	0.5789	0.4319	0.3776
	DAOP Be(3,1)	770	0.1995	0.5775	0.4279	0.3781

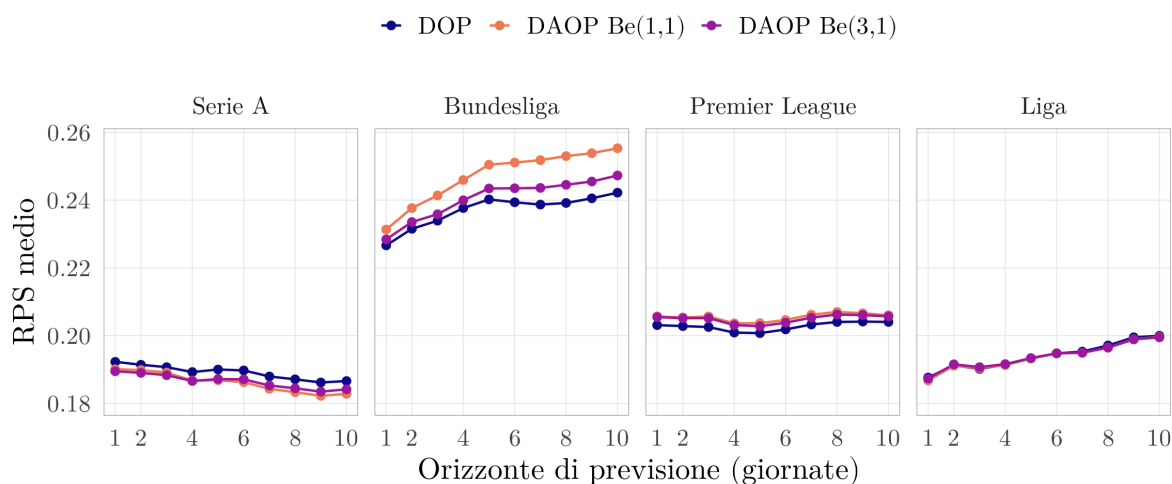


FIGURA 3.14: RPS medio in funzione dell'orizzonte di previsione, per campionato e specificazione.

Il quadro che emerge da questo secondo schema è più regolare di quello del confronto a blocco fisso. Come si osserva nella Figura 3.14, la specificazione preferibile in termini di RPS non varia con l'orizzonte ma varia con il campionato. In Bundesliga e nella Premier League DOP risulta il modello migliore, mentre in Serie A le specificazioni adattive sono leggermente preferibili. Per la Liga si ottengono risultati invece pressoché equivalenti. La regolarità riguarda dunque la famiglia di appartenenza e non la singola

calibrazione, poiché entro il modello DAOP il primato passa dall'una all'altra al variare dell'orizzonte.

Le altre tre metriche confermano questa lettura. Considerando la media sui dieci orizzonti, Brier *score* e pseudo- R^2 indicano DOP come miglior modello per Bundesliga e Premier League e una delle due calibrazioni di DAOP per Serie A e Liga, in accordo con il RPS. L'unica discordanza riguarda l'ACP nella Premier League, rispetto al quale il valore più elevato spetta a DAOP con priori $\text{Be}(1, 1)$ anziché a DOP. Considerando le metriche medie su tutti gli orizzonti, il campionato tedesco è l'unico nel quale la scelta della specificazione produce una differenza di rilievo, ed è anche quello in cui tutti i modelli ottengono i punteggi peggiori.

Una graduatoria complessiva fra le tre specificazioni non risulta tuttavia robusta. Aggregando i quattro campionati, il modello DOP risulta la specificazione migliore con un RPS medio di 0.2057, contro 0.2065 del modello DAOP $\text{Be}(3, 1)$ e 0.2081 del modello DAOP $\text{Be}(1, 1)$; escludendo la Bundesliga l'ordinamento si inverte, con entrambe le calibrazioni del modello DAOP a 0.1951 e il modello DOP a 0.1953. Poiché non vi è ragione di preferire l'una o l'altra aggregazione, la conclusione difendibile non è che una specificazione domini l'altra, bensì che la convenienza della varianza squadra-specifica sembri dipendere dalle caratteristiche del campionato nel periodo considerato.

Si analizza ora in particolare la Bundesliga, dove si presentano sia il livello peggiore delle metriche sia il divario più ampio fra specificazioni. Una spiegazione di questi risultati si può trovare nei dati. Infatti, nella finestra di valutazione, il vantaggio di giocare in casa risulta rovesciato. A fronte di una quota storica di vittorie interne pari al 43.7% sull'intero campione, nelle giornate valutate questo valore scende al 27.0%, mentre le vittorie in trasferta salgono dal 31.0 al 42.9%. Facendo riferimento all'analisi del paragrafo 3.6.1 questo fenomeno penalizza in questo contesto tutti i modelli, in particolare le specificazioni adattive.

Un ultimo elemento riguarda l'effetto dell'orizzonte, che in Serie A appare contro-intuitivo, poiché il RPS diminuisce all'allungarsi della finestra, passando da 0.1923 a 0.1866. L'andamento è attribuibile a una caratteristica dello schema descritto nel paragrafo 3.2, nel quale le giornate poste agli estremi dell'intervallo di valutazione entrano in un numero minore di finestre e ricevono pertanto un peso inferiore, tanto più ridotto quanto maggiore è H . Pesare meno le ultime giornate di campionato non costituisce peraltro necessariamente un difetto, dal momento che in esse subentrano dinamiche estranee al rendimento, legate all'assenza di obiettivi di classifica per numerose squadre. Attribuendo a ciascuna giornata il medesimo peso l'andamento discendente scompare, e in tutti e quattro i campionati si osserva un lieve peggioramento nelle previsioni.

3.7 Diagnostica di convergenza

Le stime riportate nei paragrafi precedenti sono ottenute da quattro catene indipendenti di 10 000 iterazioni ciascuna, delle quali le prime 1 500 scartate come periodo di transizione, con inizializzazioni disperse. La convergenza è valutata mediante il fattore di riduzione di scala potenziale $\hat{R}$ e l'*effective sample size*, introdotti nel paragrafo 2.3. Le diagnostiche complete, distinte per campionato, scenario, modello e blocco di parametri, sono riportate nell'Appendice A.2.

L'effetto casa, la varianza dell'errore e le varianze di innovazione rispettano la soglia $\hat{R} \leq 1.01$ in tutte le combinazioni considerate. Gli scostamenti si concentrano nelle forze latenti $\beta_{k,t}$ e sono riconducibili a due cause distinte, entrambe attese.

La prima riguarda le squadre prive di partite nell'insieme di stima. Gli indici di squadra sono definiti sull'unione degli organici delle cinque stagioni, cosicché in uno scenario che si arresta prima dell'ultima stagione alcune squadre neopromosse non compaiono in alcun incontro osservato. Le loro forze non ricevono informazione campionaria e restano determinate dalla sola distribuzione a priori e dalla dinamica di *random walk*. Si tratta di 35 celle per ciascuna delle tre specificazioni, dunque di 105 parametri su 8 268 complessivamente stimati nei quattro campionati e nei tre scenari, pari all'1.3%, con $\hat{R}$ mediano pari a 1.098 e *bulk*-ESS mediano di 33 estrazioni contro le 2 039 delle rimanenti. Nessuna di esse raggiunge la soglia $\hat{R} \leq 1.01$, in alcuna delle tre specificazioni.

La seconda causa riguarda la calibrazione della priori sul peso della mistura nel modello DAOP. La frazione di parametri entro la soglia $\hat{R} \leq 1.01$ è pari al 97.0 per cento sotto DOP e al 96.4 per cento sotto DAOP con priori $\text{Be}(1, 1)$, ma scende all'89.4 per cento sotto $\text{Be}(3, 1)$, mentre il *bulk*-ESS mediano passa da 3 994 a 1 329. Il comportamento è coerente con la struttura della mistura, poiché una distribuzione a priori più concentrata sulla componente *spike* riduce la frequenza con cui l'indicatrice di componente cambia stato, e con essa l'efficienza del campionamento.

La Figura 3.15 riassume entrambi gli aspetti considerati. Nel complesso dunque è stata ottenuta una convergenza ottimale delle catene nella quasi totalità dei casi.

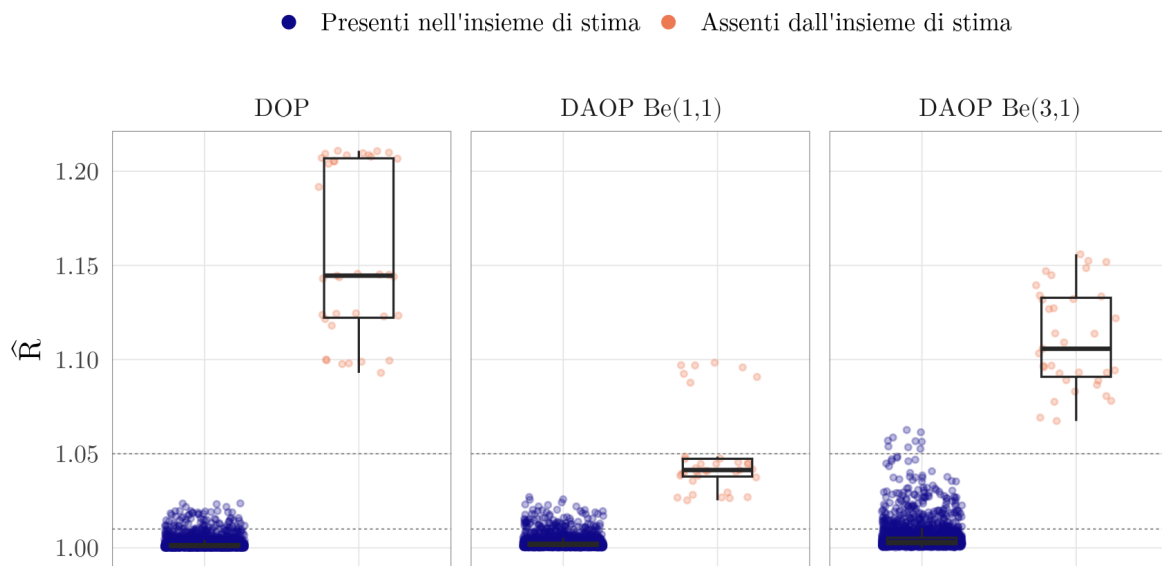


FIGURA 3.15: Distribuzione di $\hat{R}$ per le celle di $\beta_{k,t}$, per le specificazioni dinamiche, distinguendo i valori relativi a parametri di squadre assenti dall'insieme di stima. Le linee tratteggiate indicano le soglie 1.01 e 1.05.

Conclusioni

In questa tesi è stato proposto un modello ordinale dinamico con un approccio bayesiano, e ne è stata valutata sia la capacità predittiva che l'interpretabilità in due diverse specificazioni.

Dal punto di vista delle previsioni, è emersa una buona competitività del modello ordinale con i modelli *goal-based* proposti in Macrì-Demartino et al. (2026), e in alcuni casi sono state ottenute delle prestazioni migliori. Il risultato ottenuto appare interessante sia per la minore quantità di informazione considerata, che per la semplicità della struttura del modello e dell'algoritmo proposto. L'algoritmo di *Gibbs sampling* è stato infatti implementato con soli passi coniugati, e questo si è tradotto in un guadagno considerevole in termini di efficienza computazionale rispetto ai modelli concorrenti stimati tramite metodi HMC.

Per quanto riguarda l'interpretabilità, è stato possibile ricavare un ordinamento dinamico delle squadre, dal quale sono emerse buone capacità del modello di valutare la forza di ogni squadra, ad esempio nell'identificare le squadre più deboli in Serie A e che poi sono di fatto retrocesse. Per quanto concerne il confronto tra le specificazioni proposte per la varianza di evoluzione sono stati indagati dei casi particolari, in corrispondenza di eventi critici come un cambio di guida tecnica, nei quali è stato possibile apprezzare una diversa valutazione della dinamica nei parametri di forza. Proprio per il motivo per cui era stata proposta da Macrì-Demartino et al. (2026), la specificazione adattiva estesa al modello ordinale ha mostrato il comportamento atteso, riconoscendo dei momenti critici e di conseguenza agendo sulla dinamica dei parametri in maniera netta, come mostrato nei casi studio esaminati. Non sono stati ottenuti tuttavia risultati che portano a preferire una delle due specificazioni dal punto di vista predittivo in modo sistematico. L'analisi sui parametri ha comunque confermato la bontà della metodologia adattiva, che risponde alle esigenze per cui è stata adottata, permettendo inoltre sia di rendere eterogenea la varianza di innovazione che di valutare la volatilità di ogni squadra.

La metodologia proposta presenta comunque alcune criticità. La specificazione del

modello basata sulla variabile latente gaussiana e la scelta di fissare le soglie portano ad una semplificazione notevole dell'interpretazione e delle procedure di stima, ma pongono dei limiti strutturali al modello come ad esempio un limite superiore alla probabilità del pareggio, come riportato nel paragrafo 3.6.

In merito agli sviluppi futuri, si potrebbero considerare diverse proposte. La prima riguarda la distribuzione assunta per l'errore, dalla quale dipende il trattamento del pareggio. Adottando una distribuzione a code più leggere, la probabilità del pareggio negli incontri equilibrati aumenterebbe, attenuando il limite osservato in precedenza. Potrebbe essere inoltre ragionevole far dipendere la varianza dell'errore da caratteristiche osservabili dell'incontro, quali il tipo di competizione o la rilevanza della partita ai fini della classifica, rendendo il modello più flessibile proprio nella componente che ne governa la dispersione delle previsioni.

Una seconda direzione consiste nel rendere la probabilità di appartenenza alla componente diffusa della varianza di innovazione funzione di covariate osservabili, quali il movimento di mercato o l'avvicendamento nella guida tecnica. In questo modo il modello passerebbe dal riconoscere i cambiamenti di rendimento all'anticiparli sulla base delle circostanze che li determinano.

Al fine di migliorare le previsioni si potrebbero inoltre includere nel modello diverse covariate che racchiudono le informazioni sulle squadre come lo stato di forma dei giocatori, il valore della rosa, o altre statistiche utili a valutarne la capacità, oltre ad includere le quote storiche dei *bookmakers*.

Sul piano computazionale si potrebbe esplorare l'utilizzo di algoritmi di tipo *forward filtering backward sampling* al fine di campionare le traiettorie in modo congiunto, e migliorare il *mixing*.

L'ordinamento prodotto dal modello potrebbe infine essere confrontato con le graduatorie ufficiali, quali il coefficiente UEFA per i club o il *ranking* FIFA per le nazionali. Nel caso dei club europei, al fine di ottenere una graduatoria confrontabile, occorrerebbe considerare nell'insieme di stima gli incontri di molti campionati oltre alle competizioni europee in modo tale da ricondurre i parametri di ogni squadra sulla medesima scala.

Appendice

A.1 Confronto tra schemi di campionamento

La Tabella A.1 confronta il costo computazionale e l'*effective sample size* (ESS) delle due varianti di aggiornamento di β_t (paragrafo 2.4), per i modelli DOP (varianza di evoluzione comune) e DAOP (*spike-and-slab* squadra-specifico), sullo stesso scenario (Serie A 23/24 II, $n = 1310$ partite nell'insieme di stima, 10000 iterazioni con 1500 di *burn-in*, catena singola).

TABELLA A.1: Confronto del costo computazionale ed ESS tra campionamento a blocchi e univariato.

Modello	Campionamento	Tempo (s)	ESS mediano	ESS/s
DOP	a blocchi	14.17	716.9	50.59
DOP	univariato	21.38	710.3	33.22
DAOP	a blocchi	45.39	559.1	12.32
DAOP	univariato	53.20	537.2	10.10

A.2 Diagnostiche di convergenza complete

La Tabella A.2 riporta le diagnostiche di convergenza per ciascuna combinazione di campionato, scenario, modello e blocco di parametri, su quattro catene indipendenti (10 000 iterazioni, 1 500 di *burn-in*). La colonna n indica il numero di parametri considerati.

TABELLA A.2: Diagnostiche di convergenza per campionato, scenario, modello e blocco di parametri.

Lega	Scenario	Modello	Parametri	n	$\hat{R}$		%	$bulk$ -ESS	
					medio	max		medio	min
Serie A	23/24 II	SOP	α	1	1.000	1.000	100.0	15 081	15 081
	23/24 II	SOP	β_k	28	1.000	1.001	100.0	19 226	10 551
	23/24 II	SOP	σ_ε^2	1	1.003	1.003	100.0	2 672	2 672
	23/24 II	DOP	α_t	7	1.000	1.000	100.0	17 309	14 837
	23/24 II	DOP	$\beta_{k,t}$	196	1.006	1.146	96.4	4 190	19
	23/24 II	DOP	σ_ε^2	1	1.001	1.001	100.0	2 044	2 044
	23/24 II	DOP	σ_t^2	6	1.001	1.002	100.0	3 738	2 989
	23/24 II	DAOP Be(1,1)	α_t	7	1.000	1.000	100.0	16 358	13 278
	23/24 II	DAOP Be(1,1)	$\beta_{k,t}$	196	1.003	1.045	96.4	2 761	120
	23/24 II	DAOP Be(1,1)	σ_ε^2	1	1.001	1.001	100.0	1 524	1 524
	23/24 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	168	1.001	1.002	100.0	6 299	2 490
	23/24 II	DAOP Be(1,1)	$\xi_{k,t}$	168	1.001	1.003	100.0	4 608	1 962
	23/24 II	DAOP Be(1,1)	p_k	28	1.001	1.002	100.0	3 063	2 055
	23/24 II	DAOP Be(3,1)	α_t	7	1.000	1.000	100.0	16 558	13 698
	23/24 II	DAOP Be(3,1)	$\beta_{k,t}$	196	1.007	1.087	85.7	1 431	39
	23/24 II	DAOP Be(3,1)	σ_ε^2	1	1.003	1.003	100.0	1 120	1 120
	23/24 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	168	1.001	1.005	100.0	8 377	1 353
	23/24 II	DAOP Be(3,1)	$\xi_{k,t}$	168	1.001	1.007	100.0	5 225	1 020
	23/24 II	DAOP Be(3,1)	p_k	28	1.001	1.007	100.0	4 026	1 366
	24/25 II	SOP	α	1	1.000	1.000	100.0	16 101	16 101
	24/25 II	SOP	β_k	28	1.000	1.000	100.0	16 887	8 298
	24/25 II	SOP	σ_ε^2	1	1.001	1.001	100.0	2 800	2 800
	24/25 II	DOP	α_t	9	1.000	1.000	100.0	17 578	15 735
	24/25 II	DOP	$\beta_{k,t}$	252	1.002	1.015	94.8	3 894	284
	24/25 II	DOP	σ_ε^2	1	1.002	1.002	100.0	1 926	1 926
	24/25 II	DOP	σ_t^2	8	1.001	1.003	100.0	3 835	2 476
	24/25 II	DAOP Be(1,1)	α_t	9	1.000	1.000	100.0	16 765	14 470
	24/25 II	DAOP Be(1,1)	$\beta_{k,t}$	252	1.003	1.016	96.4	2 268	389
	24/25 II	DAOP Be(1,1)	σ_ε^2	1	1.004	1.004	100.0	1 524	1 524
	24/25 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	224	1.001	1.003	100.0	5 656	2 091
	24/25 II	DAOP Be(1,1)	$\xi_{k,t}$	224	1.001	1.004	100.0	3 964	1 739
	24/25 II	DAOP Be(1,1)	p_k	28	1.002	1.004	100.0	2 233	1 506
	24/25 II	DAOP Be(3,1)	α_t	9	1.000	1.001	100.0	17 063	14 303
	24/25 II	DAOP Be(3,1)	$\beta_{k,t}$	252	1.004	1.026	93.3	1 313	139
	24/25 II	DAOP Be(3,1)	σ_ε^2	1	1.004	1.004	100.0	1 374	1 374
	24/25 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	224	1.001	1.003	100.0	8 448	1 444
	24/25 II	DAOP Be(3,1)	$\xi_{k,t}$	224	1.001	1.004	100.0	5 036	1 091

continua nella pagina seguente

... segue dalla pagina precedente

Lega	Scenario	Modello	Parametri	n	medio	max	≤ 1.01	medio	min
Bundesliga	24/25 II	DAOP Be(3,1)	p_k	28	1.001	1.004	100.0	3 341	1 108
	ultima	SOP	α	1	1.000	1.000	100.0	16 192	16 192
	ultima	SOP	β_k	28	1.000	1.000	100.0	16 916	8 604
	ultima	SOP	σ_ε^2	1	1.001	1.001	100.0	3 068	3 068
	ultima	DOP	α_t	10	1.000	1.000	100.0	17 553	16 373
	ultima	DOP	$\beta_{k,t}$	280	1.002	1.019	96.4	4 062	321
	ultima	DOP	σ_ε^2	1	1.001	1.001	100.0	2 286	2 286
	ultima	DOP	σ_t^2	9	1.001	1.002	100.0	3 801	2 892
	ultima	DAOP Be(1,1)	α_t	10	1.000	1.000	100.0	17 038	14 569
	ultima	DAOP Be(1,1)	$\beta_{k,t}$	280	1.003	1.027	97.1	1 899	304
	ultima	DAOP Be(1,1)	σ_ε^2	1	1.002	1.002	100.0	1 526	1 526
	ultima	DAOP Be(1,1)	$\sigma_{k,t}^2$	252	1.001	1.004	100.0	5 666	2 276
	ultima	DAOP Be(1,1)	$\xi_{k,t}$	252	1.001	1.005	100.0	3 874	1 807
	ultima	DAOP Be(1,1)	p_k	28	1.002	1.004	100.0	2 081	1 221
	ultima	DAOP Be(3,1)	α_t	10	1.000	1.000	100.0	17 233	15 736
	ultima	DAOP Be(3,1)	$\beta_{k,t}$	280	1.006	1.057	87.1	1 241	77
	ultima	DAOP Be(3,1)	σ_ε^2	1	1.001	1.001	100.0	1 354	1 354
	ultima	DAOP Be(3,1)	$\sigma_{k,t}^2$	252	1.001	1.006	100.0	8 760	1 251
	ultima	DAOP Be(3,1)	$\xi_{k,t}$	252	1.001	1.007	100.0	5 107	922
	ultima	DAOP Be(3,1)	p_k	28	1.002	1.005	100.0	3 095	818
	23/24 II	SOP	α	1	1.000	1.000	100.0	10 426	10 426
	23/24 II	SOP	β_k	24	1.000	1.001	100.0	21 803	10 401
	23/24 II	SOP	σ_ε^2	1	1.001	1.001	100.0	2 239	2 239
	23/24 II	DOP	α_t	7	1.000	1.000	100.0	15 502	12 826
	23/24 II	DOP	$\beta_{k,t}$	168	1.015	1.210	88.1	3 569	15
	23/24 II	DOP	σ_ε^2	1	1.004	1.004	100.0	1 828	1 828
	23/24 II	DOP	σ_t^2	6	1.001	1.002	100.0	4 507	1 746
	23/24 II	DAOP Be(1,1)	α_t	7	1.000	1.000	100.0	14 657	12 281
	23/24 II	DAOP Be(1,1)	$\beta_{k,t}$	168	1.004	1.048	91.7	3 081	80
	23/24 II	DAOP Be(1,1)	σ_ε^2	1	1.002	1.002	100.0	1 244	1 244
	23/24 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	144	1.001	1.002	100.0	6 841	3 563
	23/24 II	DAOP Be(1,1)	$\xi_{k,t}$	144	1.001	1.002	100.0	5 005	2 709
	23/24 II	DAOP Be(1,1)	p_k	24	1.001	1.003	100.0	3 304	1 983
	23/24 II	DAOP Be(3,1)	α_t	7	1.000	1.000	100.0	15 683	13 415
	23/24 II	DAOP Be(3,1)	$\beta_{k,t}$	168	1.014	1.156	84.5	1 569	23
	23/24 II	DAOP Be(3,1)	σ_ε^2	1	1.001	1.001	100.0	1 262	1 262
	23/24 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	144	1.001	1.004	100.0	8 671	1 671
	23/24 II	DAOP Be(3,1)	$\xi_{k,t}$	144	1.001	1.005	100.0	5 526	1 282
	23/24 II	DAOP Be(3,1)	p_k	24	1.001	1.003	100.0	4 307	2 457
	24/25 II	SOP	α	1	1.000	1.000	100.0	11 066	11 066

continua nella pagina seguente

... segue dalla pagina precedente

Lega	Scenario	Modello	Parametri	n	medio	max	≤ 1.01	medio	min
	24/25 II	SOP	β_k	24	1.000	1.000	100.0	16 597	7 870
	24/25 II	SOP	σ_ε^2	1	1.002	1.002	100.0	2 373	2 373
	24/25 II	DOP	α_t	9	1.000	1.001	100.0	16 362	12 815
	24/25 II	DOP	$\beta_{k,t}$	216	1.004	1.020	87.0	2 699	208
	24/25 II	DOP	σ_ε^2	1	1.004	1.004	100.0	1 838	1 838
	24/25 II	DOP	σ_t^2	8	1.001	1.005	100.0	3 946	1 430
	24/25 II	DAOP Be(1,1)	α_t	9	1.000	1.000	100.0	15 778	12 720
	24/25 II	DAOP Be(1,1)	$\beta_{k,t}$	216	1.003	1.015	94.4	2 221	359
	24/25 II	DAOP Be(1,1)	σ_ε^2	1	1.005	1.005	100.0	1 470	1 470
	24/25 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	192	1.001	1.003	100.0	6 226	2 758
	24/25 II	DAOP Be(1,1)	$\xi_{k,t}$	192	1.001	1.003	100.0	4 473	2 401
	24/25 II	DAOP Be(1,1)	p_k	24	1.002	1.003	100.0	2 562	1 808
	24/25 II	DAOP Be(3,1)	α_t	9	1.000	1.000	100.0	16 077	12 845
	24/25 II	DAOP Be(3,1)	$\beta_{k,t}$	216	1.008	1.063	87.0	1 132	67
	24/25 II	DAOP Be(3,1)	σ_ε^2	1	1.006	1.006	100.0	1 067	1 067
	24/25 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	192	1.001	1.004	100.0	8 204	1 767
	24/25 II	DAOP Be(3,1)	$\xi_{k,t}$	192	1.001	1.005	100.0	5 026	1 247
	24/25 II	DAOP Be(3,1)	p_k	24	1.002	1.003	100.0	3 245	1 248
	ultima	SOP	α	1	1.000	1.000	100.0	12 161	12 161
	ultima	SOP	β_k	24	1.000	1.000	100.0	17 138	8 447
	ultima	SOP	σ_ε^2	1	1.001	1.001	100.0	2 536	2 536
	ultima	DOP	α_t	10	1.000	1.000	100.0	17 036	14 490
	ultima	DOP	$\beta_{k,t}$	240	1.003	1.024	95.4	2 970	282
	ultima	DOP	σ_ε^2	1	1.002	1.002	100.0	1 673	1 673
	ultima	DOP	σ_t^2	9	1.001	1.004	100.0	4 114	1 639
	ultima	DAOP Be(1,1)	α_t	10	1.000	1.001	100.0	16 145	13 550
	ultima	DAOP Be(1,1)	$\beta_{k,t}$	240	1.003	1.020	94.6	1 946	271
	ultima	DAOP Be(1,1)	σ_ε^2	1	1.001	1.001	100.0	1 265	1 265
	ultima	DAOP Be(1,1)	$\sigma_{k,t}^2$	216	1.001	1.003	100.0	5 733	2 179
	ultima	DAOP Be(1,1)	$\xi_{k,t}$	216	1.001	1.003	100.0	4 045	1 909
	ultima	DAOP Be(1,1)	p_k	24	1.002	1.004	100.0	2 172	1 271
	ultima	DAOP Be(3,1)	α_t	10	1.000	1.000	100.0	16 775	13 957
	ultima	DAOP Be(3,1)	$\beta_{k,t}$	240	1.006	1.029	82.9	1 091	157
	ultima	DAOP Be(3,1)	σ_ε^2	1	1.002	1.002	100.0	1 180	1 180
	ultima	DAOP Be(3,1)	$\sigma_{k,t}^2$	216	1.001	1.004	100.0	8 119	1 400
	ultima	DAOP Be(3,1)	$\xi_{k,t}$	216	1.001	1.005	100.0	4 760	953
	ultima	DAOP Be(3,1)	p_k	24	1.002	1.008	100.0	2 849	1 023
Premier League	23/24 II	SOP	α	1	1.000	1.000	100.0	13 715	13 715
	23/24 II	SOP	β_k	27	1.000	1.001	100.0	17 895	6 878
	23/24 II	SOP	σ_ε^2	1	1.002	1.002	100.0	2 038	2 038

continua nella pagina seguente

... segue dalla pagina precedente

Lega	Scenario	Modello	Parametri	n	medio	max	≤ 1.01	medio	min
	23/24 II	DOP	α_t	7	1.000	1.000	100.0	16 388	14 211
	23/24 II	DOP	$\beta_{k,t}$	189	1.010	1.211	94.2	3 547	14
	23/24 II	DOP	σ_ε^2	1	1.003	1.003	100.0	1 526	1 526
	23/24 II	DOP	σ_t^2	6	1.002	1.003	100.0	2 947	2 103
	23/24 II	DAOP Be(1,1)	α_t	7	1.000	1.000	100.0	15 346	12 798
	23/24 II	DAOP Be(1,1)	$\beta_{k,t}$	189	1.003	1.041	96.3	3 212	108
	23/24 II	DAOP Be(1,1)	σ_ε^2	1	1.002	1.002	100.0	1 371	1 371
	23/24 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	162	1.001	1.002	100.0	6 555	3 282
	23/24 II	DAOP Be(1,1)	$\xi_{k,t}$	162	1.001	1.003	100.0	4 856	2 336
	23/24 II	DAOP Be(1,1)	p_k	27	1.001	1.003	100.0	3 205	2 438
	23/24 II	DAOP Be(3,1)	α_t	7	1.000	1.000	100.0	15 298	10 997
	23/24 II	DAOP Be(3,1)	$\beta_{k,t}$	189	1.008	1.114	89.4	1 455	29
	23/24 II	DAOP Be(3,1)	σ_ε^2	1	1.002	1.002	100.0	994	994
	23/24 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	162	1.001	1.003	100.0	7 328	2 403
	23/24 II	DAOP Be(3,1)	$\xi_{k,t}$	162	1.001	1.005	100.0	4 814	1 746
	23/24 II	DAOP Be(3,1)	p_k	27	1.001	1.004	100.0	3 667	1 901
	24/25 II	SOP	α	1	1.000	1.000	100.0	13 199	13 199
	24/25 II	SOP	β_k	27	1.000	1.001	100.0	16 430	6 876
	24/25 II	SOP	σ_ε^2	1	1.003	1.003	100.0	2 133	2 133
	24/25 II	DOP	α_t	9	1.000	1.001	100.0	16 489	13 983
	24/25 II	DOP	$\beta_{k,t}$	243	1.002	1.012	97.9	3 728	296
	24/25 II	DOP	σ_ε^2	1	1.003	1.003	100.0	1 310	1 310
	24/25 II	DOP	σ_t^2	8	1.001	1.003	100.0	3 119	2 029
	24/25 II	DAOP Be(1,1)	α_t	9	1.000	1.000	100.0	15 869	13 902
	24/25 II	DAOP Be(1,1)	$\beta_{k,t}$	243	1.002	1.020	95.1	2 523	367
	24/25 II	DAOP Be(1,1)	σ_ε^2	1	1.003	1.003	100.0	1 477	1 477
	24/25 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	216	1.001	1.002	100.0	6 043	2 508
	24/25 II	DAOP Be(1,1)	$\xi_{k,t}$	216	1.001	1.004	100.0	4 386	1 896
	24/25 II	DAOP Be(1,1)	p_k	27	1.001	1.003	100.0	2 493	1 733
	24/25 II	DAOP Be(3,1)	α_t	9	1.000	1.000	100.0	16 090	14 155
	24/25 II	DAOP Be(3,1)	$\beta_{k,t}$	243	1.005	1.038	87.7	1 338	140
	24/25 II	DAOP Be(3,1)	σ_ε^2	1	1.003	1.003	100.0	948	948
	24/25 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	216	1.001	1.004	100.0	7 384	2 131
	24/25 II	DAOP Be(3,1)	$\xi_{k,t}$	216	1.001	1.005	100.0	4 639	1 503
	24/25 II	DAOP Be(3,1)	p_k	27	1.002	1.004	100.0	3 063	1 236
	ultima	SOP	α	1	1.000	1.000	100.0	14 696	14 696
	ultima	SOP	β_k	27	1.000	1.001	100.0	16 082	6 645
	ultima	SOP	σ_ε^2	1	1.001	1.001	100.0	2 346	2 346
	ultima	DOP	α_t	10	1.000	1.000	100.0	16 824	14 566
	ultima	DOP	$\beta_{k,t}$	270	1.002	1.007	100.0	3 643	420

continua nella pagina seguente

... segue dalla pagina precedente

Lega	Scenario	Modello	Parametri	n	medio	max	≤ 1.01	medio	min
LaLiga	ultima	DOP	σ_ε^2	1	1.002	1.002	100.0	1 548	1 548
	ultima	DOP	σ_t^2	9	1.001	1.003	100.0	3 138	1 830
	ultima	DAOP Be(1,1)	α_t	10	1.000	1.001	100.0	15 956	13 963
	ultima	DAOP Be(1,1)	$\beta_{k,t}$	270	1.002	1.014	97.4	2 351	356
	ultima	DAOP Be(1,1)	σ_ε^2	1	1.004	1.004	100.0	1 261	1 261
	ultima	DAOP Be(1,1)	$\sigma_{k,t}^2$	243	1.001	1.002	100.0	5 875	2 824
	ultima	DAOP Be(1,1)	$\xi_{k,t}$	243	1.001	1.003	100.0	4 251	1 897
	ultima	DAOP Be(1,1)	p_k	27	1.002	1.005	100.0	2 304	1 686
	ultima	DAOP Be(3,1)	α_t	10	1.000	1.001	100.0	15 726	13 478
	ultima	DAOP Be(3,1)	$\beta_{k,t}$	270	1.005	1.038	88.5	1 244	119
	ultima	DAOP Be(3,1)	σ_ε^2	1	1.003	1.003	100.0	650	650
	ultima	DAOP Be(3,1)	$\sigma_{k,t}^2$	243	1.001	1.005	100.0	6 900	1 632
	ultima	DAOP Be(3,1)	$\xi_{k,t}$	243	1.001	1.007	100.0	4 325	1 019
	ultima	DAOP Be(3,1)	p_k	27	1.002	1.007	100.0	2 538	1 097
	23/24 II	SOP	α	1	1.000	1.000	100.0	11 044	11 044
	23/24 II	SOP	β_k	27	1.000	1.000	100.0	20 685	11 812
	23/24 II	SOP	σ_ε^2	1	1.001	1.001	100.0	2 986	2 986
	23/24 II	DOP	α_t	7	1.000	1.000	100.0	16 306	13 310
	23/24 II	DOP	$\beta_{k,t}$	189	1.006	1.125	96.3	4 707	33
	23/24 II	DOP	σ_ε^2	1	1.002	1.002	100.0	1 956	1 956
	23/24 II	DOP	σ_t^2	6	1.001	1.002	100.0	4 416	1 685
	23/24 II	DAOP Be(1,1)	α_t	7	1.000	1.001	100.0	16 039	10 491
	23/24 II	DAOP Be(1,1)	$\beta_{k,t}$	189	1.005	1.098	94.7	2 835	34
	23/24 II	DAOP Be(1,1)	σ_ε^2	1	1.004	1.004	100.0	1 608	1 608
	23/24 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	162	1.001	1.003	100.0	6 530	3 014
	23/24 II	DAOP Be(1,1)	$\xi_{k,t}$	162	1.001	1.004	100.0	4 625	2 339
	23/24 II	DAOP Be(1,1)	p_k	27	1.001	1.004	100.0	3 092	1 785
	23/24 II	DAOP Be(3,1)	α_t	7	1.000	1.000	100.0	15 931	12 545
	23/24 II	DAOP Be(3,1)	$\beta_{k,t}$	189	1.007	1.134	95.8	1 829	29
	23/24 II	DAOP Be(3,1)	σ_ε^2	1	1.001	1.001	100.0	1 489	1 489
	23/24 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	162	1.000	1.003	100.0	9 734	2 549
	23/24 II	DAOP Be(3,1)	$\xi_{k,t}$	162	1.001	1.004	100.0	5 973	1 815
	23/24 II	DAOP Be(3,1)	p_k	27	1.001	1.003	100.0	4 970	2 076
	24/25 II	SOP	α	1	1.000	1.000	100.0	10 039	10 039
	24/25 II	SOP	β_k	27	1.000	1.000	100.0	17 731	8 331
	24/25 II	SOP	σ_ε^2	1	1.001	1.001	100.0	2 684	2 684
	24/25 II	DOP	α_t	9	1.000	1.001	100.0	16 397	14 577
	24/25 II	DOP	$\beta_{k,t}$	243	1.002	1.012	97.9	3 977	331
	24/25 II	DOP	σ_ε^2	1	1.003	1.003	100.0	1 921	1 921
	24/25 II	DOP	σ_t^2	8	1.001	1.003	100.0	4 337	3 000

continua nella pagina seguente

... segue dalla pagina precedente

Lega	Scenario	Modello	Parametri	n	medio	max	≤ 1.01	medio	min
	24/25 II	DAOP Be(1,1)	α_t	9	1.000	1.000	100.0	15 945	12 500
	24/25 II	DAOP Be(1,1)	$\beta_{k,t}$	243	1.003	1.020	93.4	1 990	335
	24/25 II	DAOP Be(1,1)	σ_ε^2	1	1.002	1.002	100.0	1 562	1 562
	24/25 II	DAOP Be(1,1)	$\sigma_{k,t}^2$	216	1.001	1.003	100.0	5 718	2 096
	24/25 II	DAOP Be(1,1)	$\xi_{k,t}$	216	1.001	1.004	100.0	3 847	1 540
	24/25 II	DAOP Be(1,1)	p_k	27	1.002	1.005	100.0	2 232	1 462
	24/25 II	DAOP Be(3,1)	α_t	9	1.000	1.000	100.0	16 243	11 886
	24/25 II	DAOP Be(3,1)	$\beta_{k,t}$	243	1.005	1.028	89.7	1 333	133
	24/25 II	DAOP Be(3,1)	σ_ε^2	1	1.002	1.002	100.0	1 184	1 184
	24/25 II	DAOP Be(3,1)	$\sigma_{k,t}^2$	216	1.001	1.004	100.0	9 608	1 349
	24/25 II	DAOP Be(3,1)	$\xi_{k,t}$	216	1.001	1.005	100.0	5 517	930
	24/25 II	DAOP Be(3,1)	p_k	27	1.001	1.005	100.0	3 849	1 025
ultima	SOP		α	1	1.000	1.000	100.0	12 019	12 019
ultima	SOP		β_k	27	1.000	1.001	100.0	17 853	8 930
ultima	SOP		σ_ε^2	1	1.001	1.001	100.0	2 859	2 859
ultima	DOP		α_t	10	1.000	1.000	100.0	16 927	15 709
ultima	DOP		$\beta_{k,t}$	270	1.002	1.010	100.0	4 129	410
ultima	DOP		σ_ε^2	1	1.002	1.002	100.0	1 990	1 990
ultima	DOP		σ_t^2	9	1.001	1.002	100.0	3 951	2 159
ultima	DAOP Be(1,1)		α_t	10	1.000	1.000	100.0	16 031	12 865
ultima	DAOP Be(1,1)		$\beta_{k,t}$	270	1.003	1.023	93.3	1 987	260
ultima	DAOP Be(1,1)		σ_ε^2	1	1.001	1.001	100.0	1 512	1 512
ultima	DAOP Be(1,1)		$\sigma_{k,t}^2$	243	1.001	1.004	100.0	5 647	1 878
ultima	DAOP Be(1,1)		$\xi_{k,t}$	243	1.001	1.005	100.0	3 838	1 577
ultima	DAOP Be(1,1)		p_k	27	1.002	1.008	100.0	2 083	964
ultima	DAOP Be(3,1)		α_t	10	1.000	1.001	100.0	16 229	14 025
ultima	DAOP Be(3,1)		$\beta_{k,t}$	270	1.005	1.032	87.8	1 263	138
ultima	DAOP Be(3,1)		σ_ε^2	1	1.005	1.005	100.0	1 344	1 344
ultima	DAOP Be(3,1)		$\sigma_{k,t}^2$	243	1.001	1.004	100.0	9 192	1 565
ultima	DAOP Be(3,1)		$\xi_{k,t}$	243	1.001	1.006	100.0	5 344	1 120
ultima	DAOP Be(3,1)		p_k	27	1.001	1.009	100.0	3 411	838

A.3 Analisi di sensitività

Si riportano le figure dell'analisi di sensitività discussa nel paragrafo 3.3. Per ciascun campionato sono considerate due squadre, scelte in modo da rappresentare rispettivamente un rendimento stabile e uno mutevole. Ogni pannello riporta, per ogni periodo, la dispersione delle mediane a posteriori di $\beta_{k,t}$ sulle nove configurazioni della griglia.



FIGURA A.1: Sensitività del modello DOP agli iperparametri della priori su σ_t^2 . I rettangoli riassumono la distribuzione delle mediane a posteriori di $\beta_{k,t}$ sulle nove configurazioni ottenute incrociando $a \in \{1.5, 2, 3\}$ con $E[\sigma_t^2] \in \{0.05, 0.1, 0.2\}$. I punti grigi rappresentano le singole configurazioni e il triangolo la calibrazione adottata, $a = 2$ e $E[\sigma_t^2] = 0.1$.



FIGURA A.2: Sensitività del modello DAOP agli iperparametri della mistura *spike-and-slab*. I rettangoli riassumono la distribuzione delle mediane a posteriori di $\beta_{k,t}$ sulle nove configurazioni ottenute riscaldando di un fattore $\{0.5, 1, 2\}$ le mediane a priori delle componenti *spike* e *slab*. I punti grigi rappresentano le singole configurazioni e il triangolo la calibrazione adottata.

A.4 Evoluzione dei *rankings*

Si riportano, per i tre campionati non discussi nel paragrafo 3.5.2, i *rankings* dinamici sotto ciascuna delle tre specificazioni.

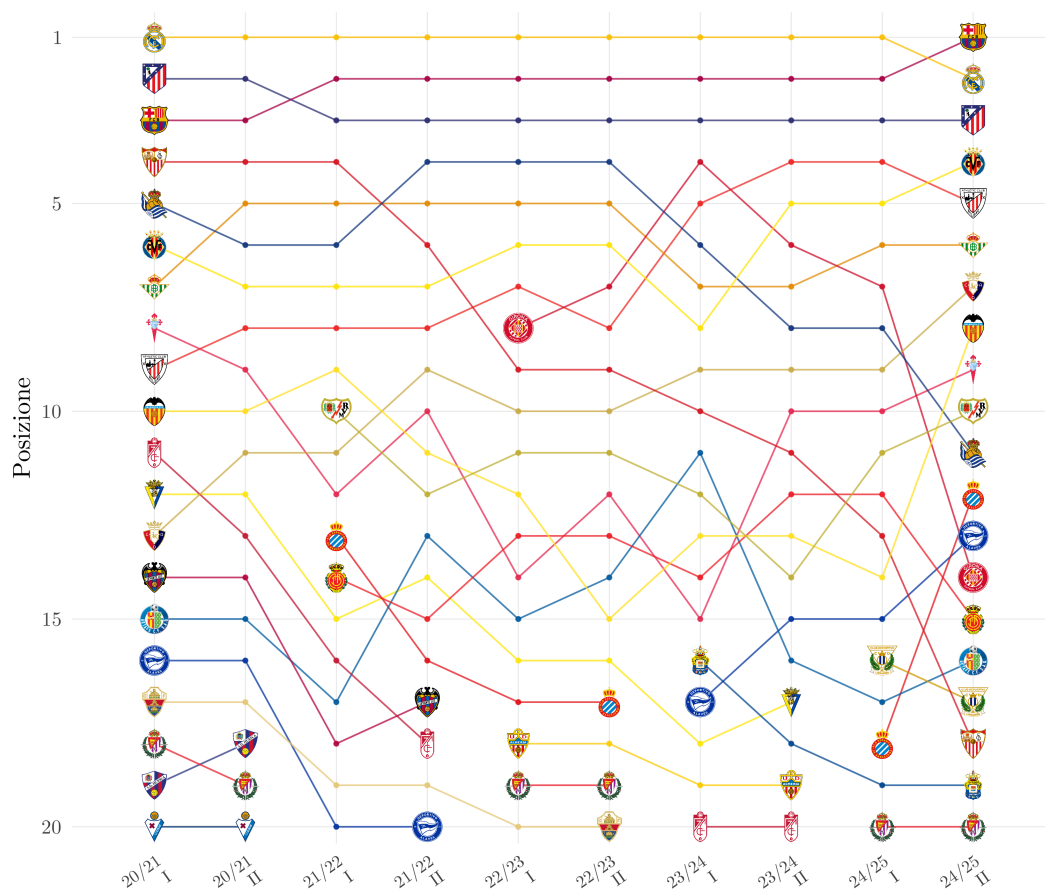


FIGURA A.3: Evoluzione della posizione nella classifica latente, Liga, specificazione DOP.

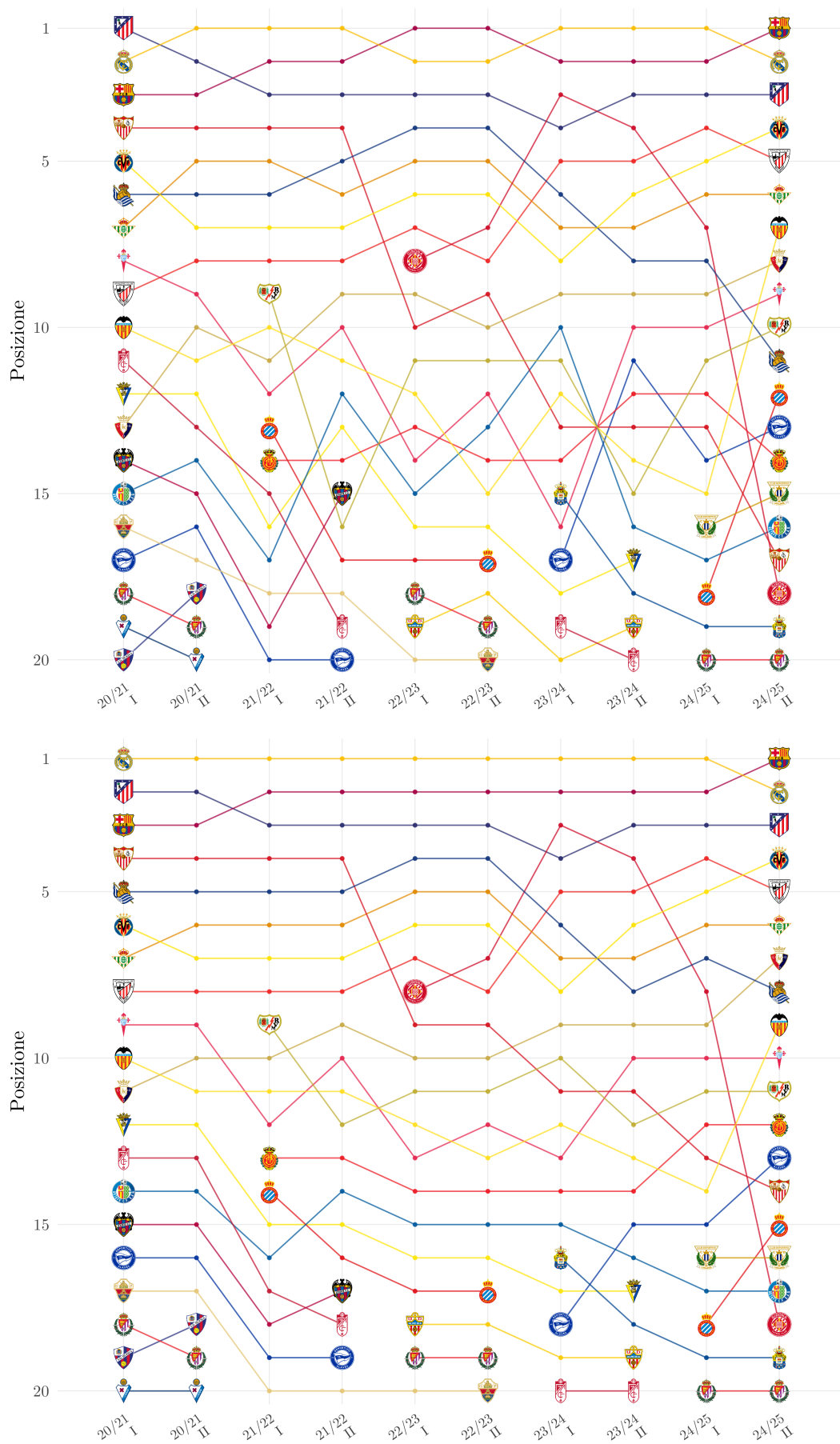


FIGURA A.4: Evoluzione della posizione nella classifica latente, Liga, specificazione DAOP, priori $Be(1,1)$ (in alto) e $Be(3,1)$ (in basso).

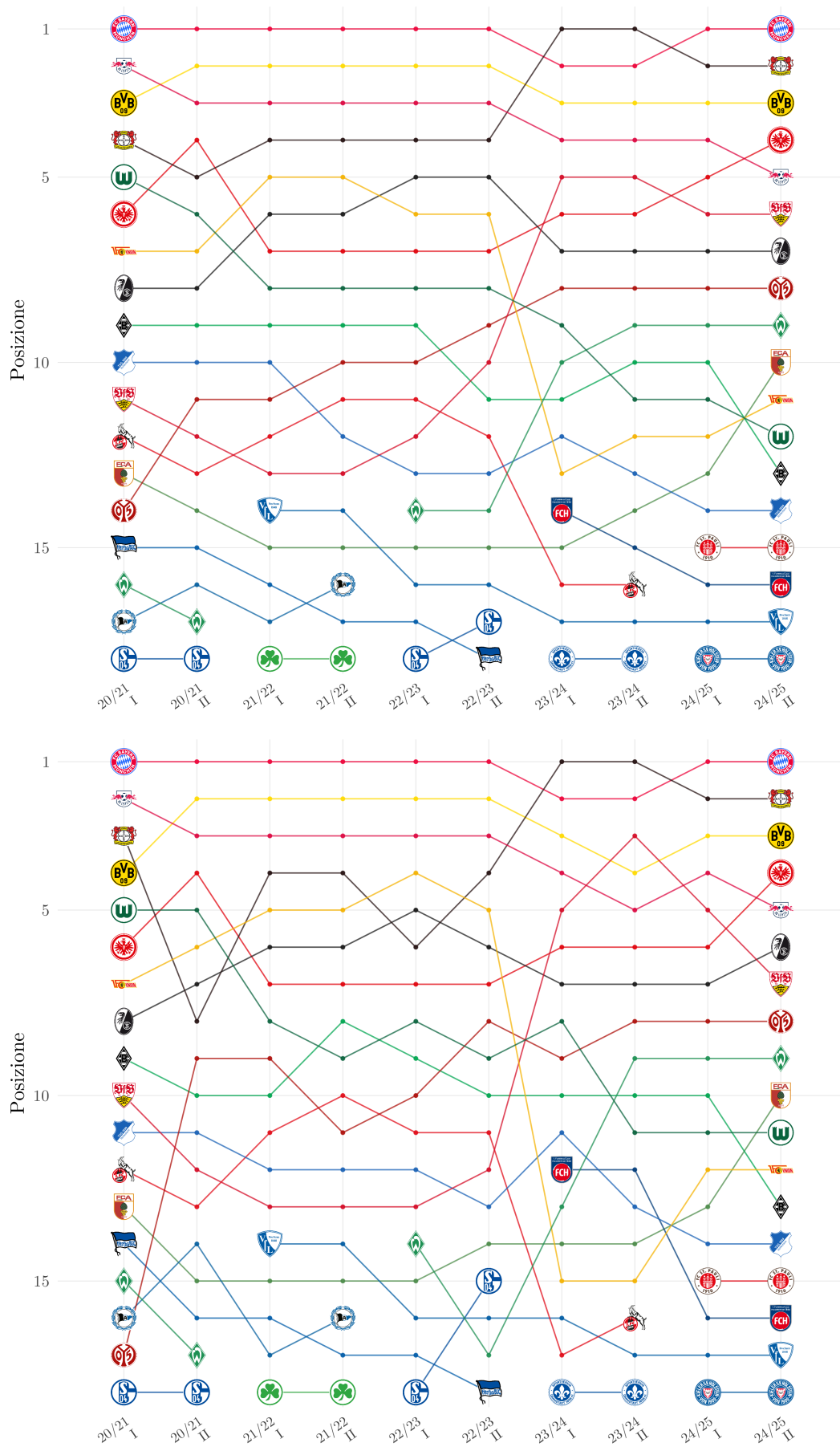


FIGURA A.5: Evoluzione della posizione nella classifica latente, Bundesliga, specificazione DOP (in alto) e DAOP con priori $Be(1,1)$ (in basso).

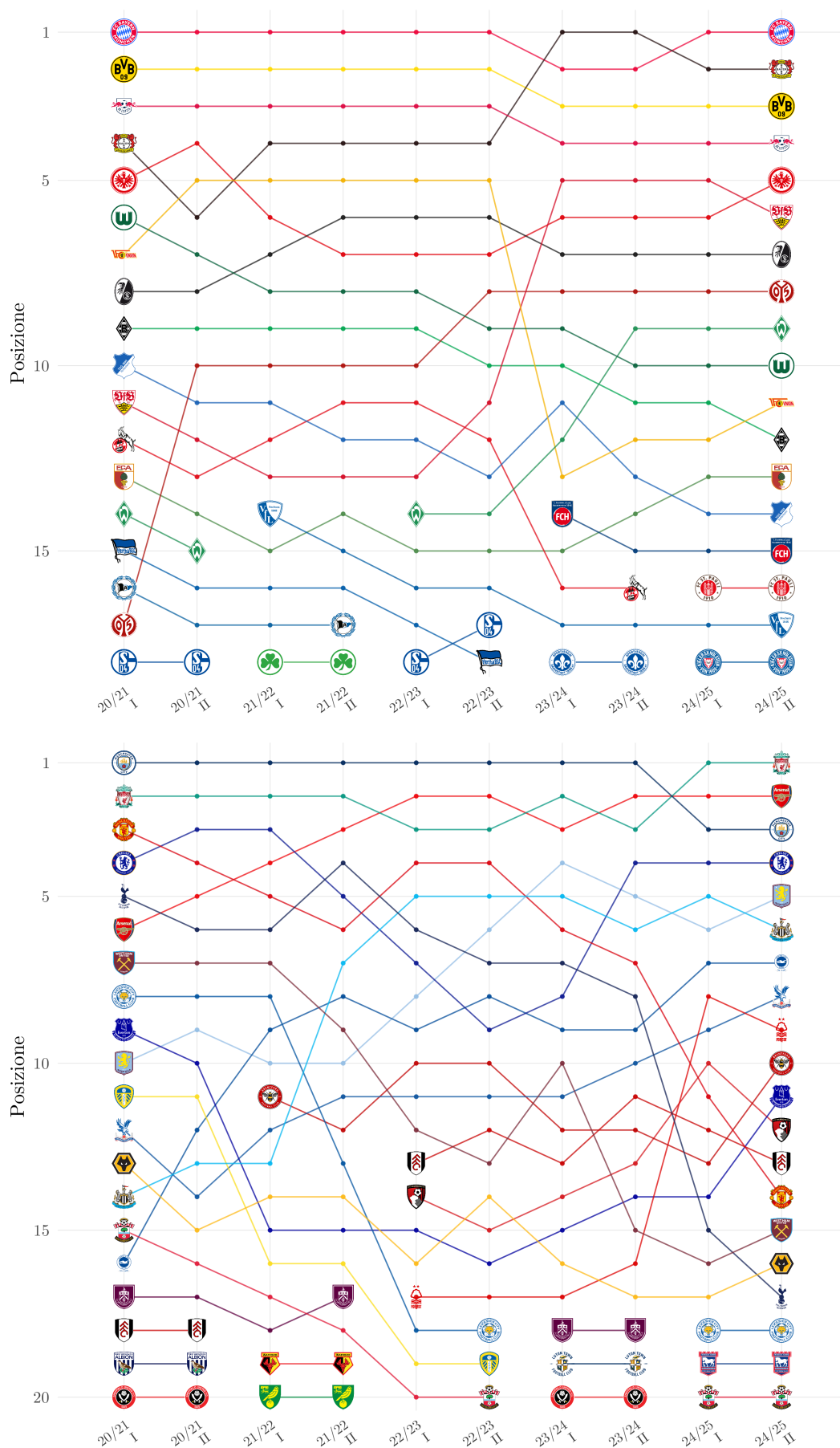


FIGURA A.6: Evoluzione della posizione nella classifica latente, Bundesliga, specificazione DAOP con priori $Be(3, 1)$ (in alto) e Premier League, specificazione DOP (in basso).

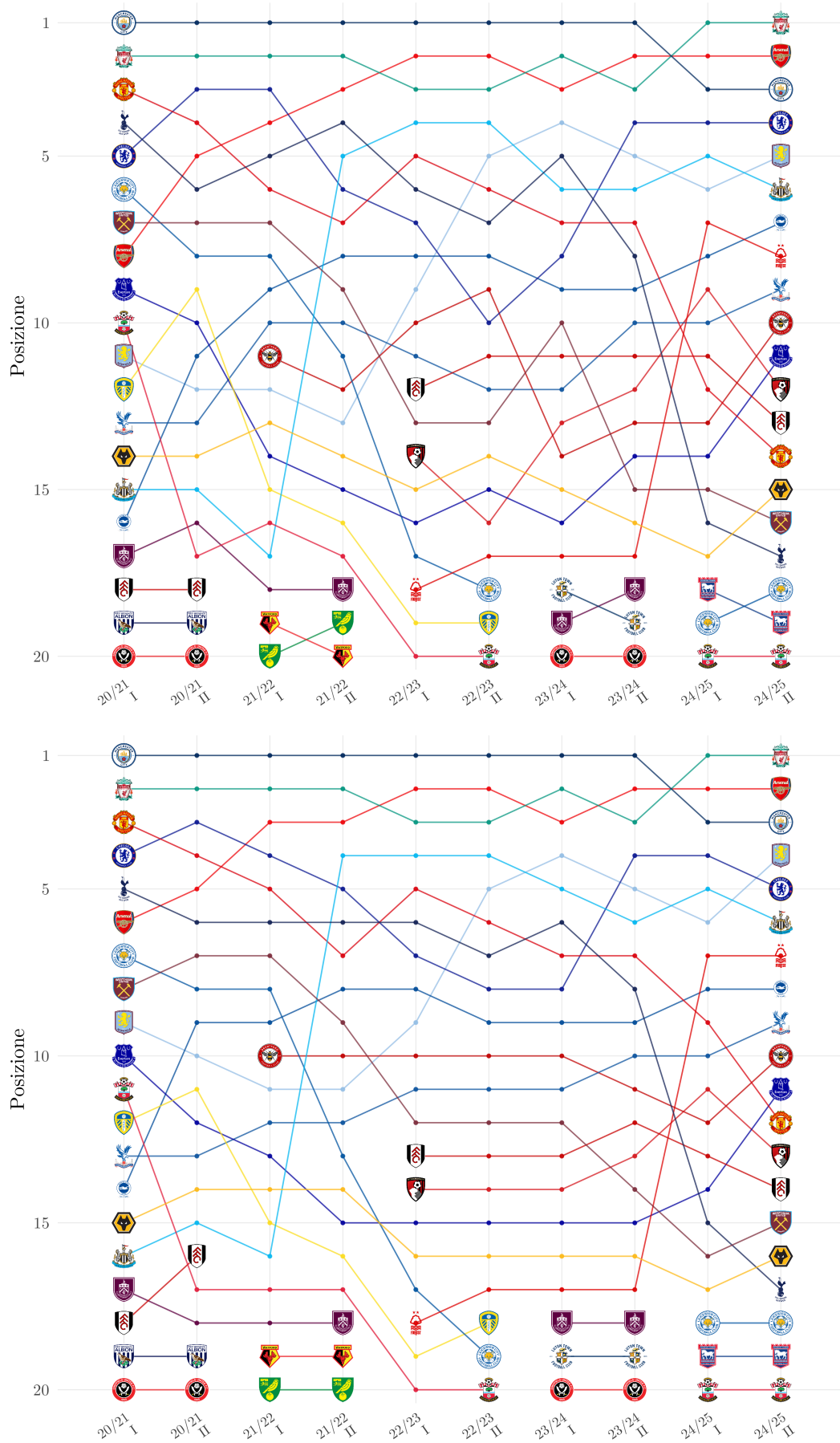


FIGURA A.7: Evoluzione della posizione nella classifica latente, Premier League, specificazione DAOP: priori $\text{Be}(1,1)$ (in alto) e $\text{Be}(3,1)$ (in basso).

A.5 Pesi della componente *slab*

Si riporta, per i tre campionati non discussi nel paragrafo 3.5.3, la distribuzione a posteriori del peso della componente *slab*, $1 - p_k$, per le squadre presenti in almeno cinque periodi.

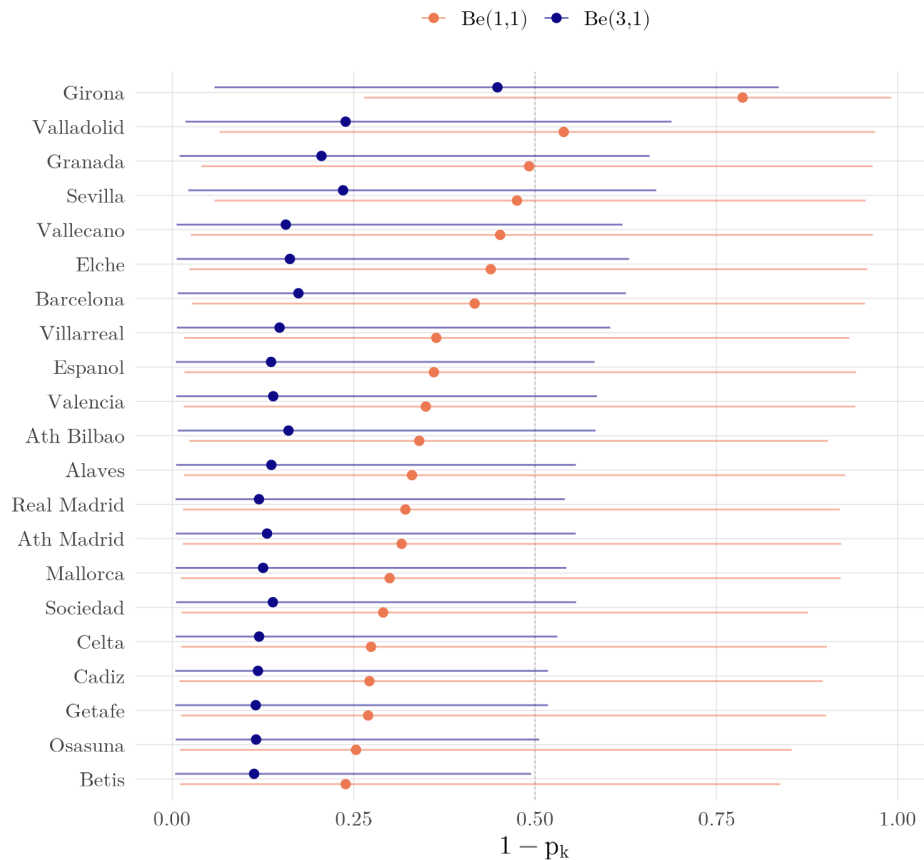


FIGURA A.8: Distribuzione a posteriori del peso della componente *slab*, $1 - p_k$, per le squadre della Liga, sotto le due calibrazioni della priori su p_k . Il punto rappresenta la mediana a posteriori, il segmento l'intervallo di credibilità al 95%. Le squadre sono ordinate secondo la mediana ottenuta sotto la calibrazione Be(1, 1).

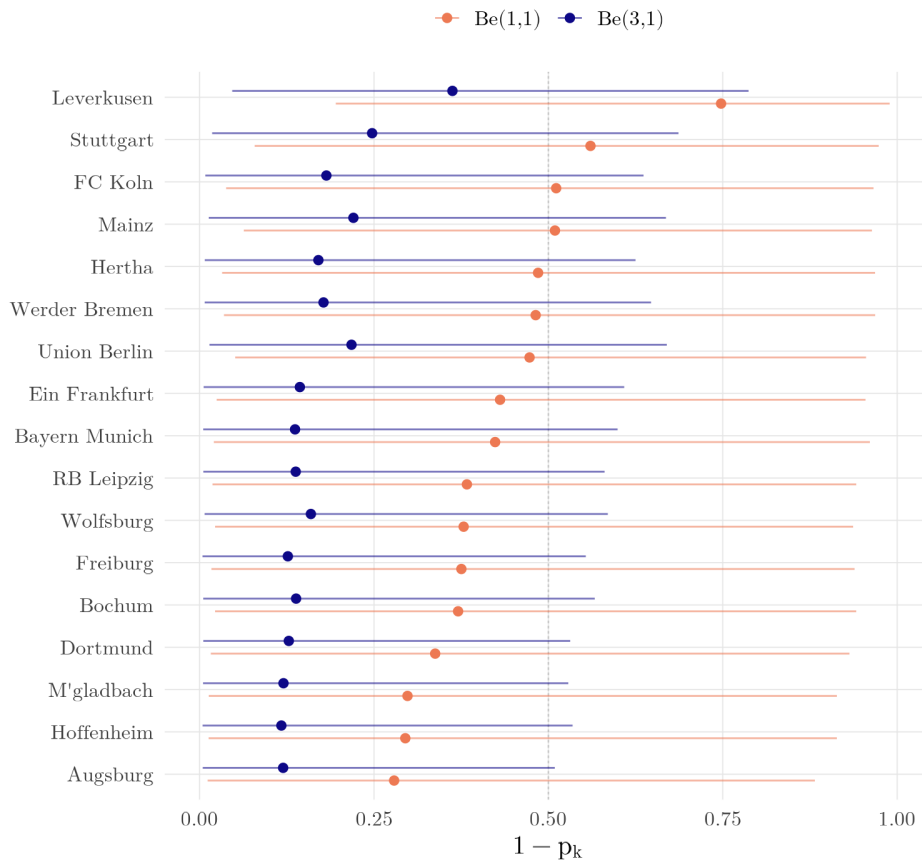


FIGURA A.9: Distribuzione a posteriori del peso della componente *slab*, $1 - p_k$, per le squadre della Bundesliga, sotto le due calibrazioni della priori su p_k . Il punto rappresenta la mediana a posteriori, il segmento l'intervallo di credibilità al 95%. Le squadre sono ordinate secondo la mediana ottenuta sotto la calibrazione Be(1, 1).

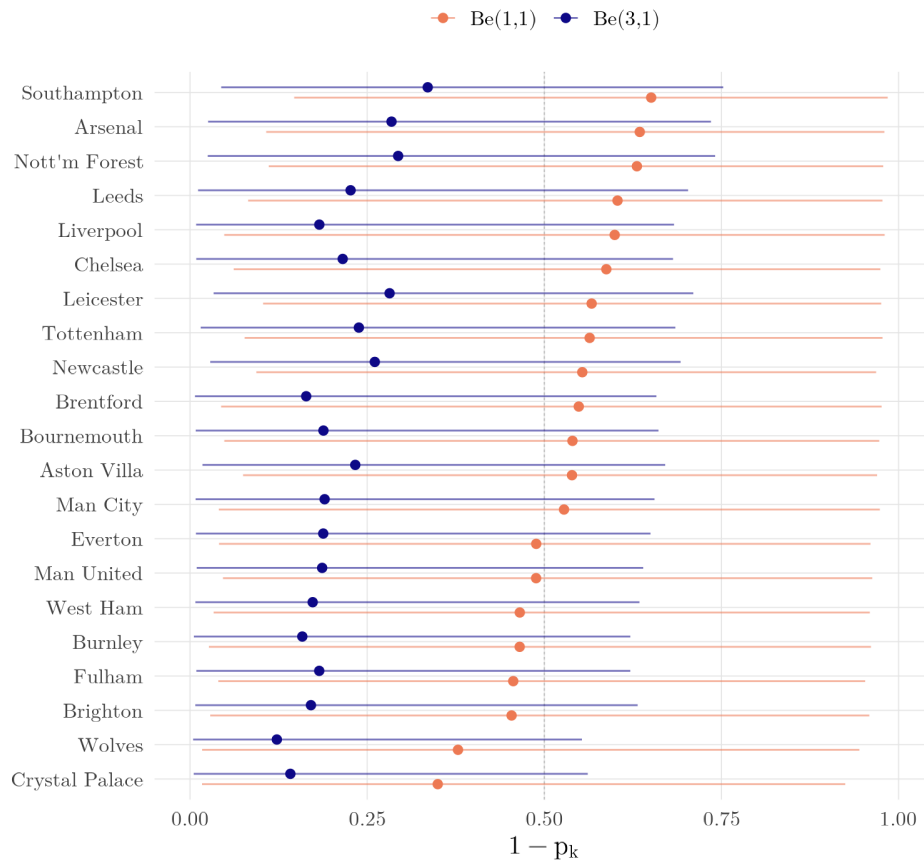


FIGURA A.10: Distribuzione a posteriori del peso della componente *slab*, $1 - p_k$, per le squadre della Premier League, sotto le due calibrazioni della priori su p_k . Il punto rappresenta la mediana a posteriori, il segmento l'intervallo di credibilità al 95%. Le squadre sono ordinate secondo la mediana ottenuta sotto la calibrazione $\text{Be}(1, 1)$.

A.6 Codice R

Si riportano di seguito i codici R relativi ai due algoritmi di *Gibbs sampling* descritti nel paragrafo 2.4.

```

1 library(truncnorm)
2
3 # Gamma inversa(a, b) con shape = a, rate = b, quindi E[x] = b / (a - 1).
4 r_inv_gamma <- function(n, a, b) {
5   1 / rgamma(n, shape = a, rate = b)
6 }
7
8 # Matrice di disegno del periodo t (n_t x K):
9 # X[i, k] = 1 se la squadra k gioca in casa nella partita i
10 #          -1 se gioca in trasferta
11 #          0 altrimenti
12 fun_X_t <- function(incontri, t_val, K) {
13   idx <- which(incontri$t == t_val)

```

```

14  n_t <- length(idx)
15  X    <- matrix(0, nrow = n_t, ncol = K)
16  for (i in seq_len(n_t)) {
17    X[i, incontri$h[idx[i]]] <- 1
18    X[i, incontri$a[idx[i]]] <- -1
19  }
20  X
21 }
22 # Predittore lineare eta[i] = alpha[t_i] + beta[h_i, t_i] - beta[a_i, t_i]
23 fun_eta <- function(incontri, beta, alpha) {
24   alpha[incontri$t] +
25     beta[cbind(incontri$h, incontri$t)] - beta[cbind(incontri$a, incontri$t)]
26 }
27 # Campionamento di Z[i], di legge N(eta[i], sigma2_eps),
28 # troncata all'intervallo imposto dall'esito osservato,
29 # y = -1 -> Z <= -0.5
30 # y = 0 -> -0.5 < Z <= 0.5
31 # y = 1 -> Z > 0.5
32 sample_Z <- function(incontri, beta, alpha, sigma2_eps) {
33   eta <- fun_eta(incontri, beta, alpha)
34   sd_eps <- sqrt(sigma2_eps)
35   Z <- numeric(nrow(incontri))
36   idx_osp <- incontri$y == -1
37   idx_par <- incontri$y == 0
38   idx_cas <- incontri$y == 1
39
40   if (any(idx_osp))
41     Z[idx_osp] <- rtruncnorm(sum(idx_osp), a = -Inf, b = -0.5,
42                             mean = eta[idx_osp], sd = sd_eps)
43   if (any(idx_par))
44     Z[idx_par] <- rtruncnorm(sum(idx_par), a = -0.5, b = 0.5,
45                             mean = eta[idx_par], sd = sd_eps)
46   if (any(idx_cas))
47     Z[idx_cas] <- rtruncnorm(sum(idx_cas), a = 0.5, b = Inf,
48                             mean = eta[idx_cas], sd = sd_eps)
49   Z
50 }
51 # Aggiornamento coniugato dell'effetto casa
52 sample_alpha <- function(incontri, Z, beta, sigma2_eps, sigma2_alpha,
53                          idx_t, n_t) {
54   T_periods <- length(idx_t)
55   alpha <- numeric(T_periods)
56   r_i <- Z - beta[cbind(incontri$h, incontri$t)] +
57     beta[cbind(incontri$a, incontri$t)]
58   for (t in seq_len(T_periods)) {
59     V_t <- 1 / (n_t[t] / sigma2_eps + 1 / sigma2_alpha)
60     mu_t <- V_t / sigma2_eps * sum(r_i[idx_t[[t]])]
61     alpha[t] <- rnorm(1, mean = mu_t, sd = sqrt(V_t))
62   }
63   alpha
64 }

```

CODICE A.1: Funzioni di servizio comuni ai due algoritmi.

```

1  # Gibbs sampler per la specificazione DOP
2  source("funzioni.R")
3
4  gibbs.multi.dinamico <- function(dati, hyp, niter = 10000, burnin = 2000) {
5
6    incontri <- dati$incontri
7    K <- dati$K
8    T_periods <- dati$T_periods
9    n <- nrow(incontri)
10
11    sigma2_0 <- hyp$sigma2_0
12    a_sig <- hyp$a
13    b_sig <- hyp$b
14    a_eps <- hyp$a_eps
15    b_eps <- hyp$b_eps
16    sigma2_alpha <- hyp$sigma2_alpha
17
18    beta <- matrix(0, nrow = K, ncol = T_periods)
19    sigma2_t <- rep(0.1, T_periods - 1)
20    sigma2_eps <- 1
21    alpha <- rep(0.3, T_periods)
22
23    idx_t <- lapply(seq_len(T_periods), function(t) which(incontri$t == t))
24    n_t <- vapply(idx_t, length, 1)
25    X_t_list <- lapply(seq_len(T_periods), function(t) fun_X_t(incontri, t, K))
26    XtX_list <- lapply(X_t_list, crossprod)
27
28    n_saved <- niter - burnin
29    beta_chain <- array(NA, dim = c(K, T_periods, n_saved))
30    sigma2_chain <- matrix(NA, nrow = T_periods - 1, ncol = n_saved)
31    sigma2_eps_chain <- numeric(n_saved)
32    alpha_chain <- matrix(NA, nrow = T_periods, ncol = n_saved)
33    save_idx <- 0
34
35    for (iter in seq_len(niter)) {
36
37      # 1:  $Z[i]$ 
38      Z <- sample_Z(incontri, beta, alpha, sigma2_eps)
39
40      # 2:  $\beta[, t]$  a blocchi
41      psi <- 1 / sigma2_t
42      for (t in seq_len(T_periods)) {
43        idx <- idx_t[[t]]
44        X_t <- X_t_list[[t]]
45        Z_tilde <- Z[idx] - alpha[t]
46
47        A <- XtX_list[[t]] / sigma2_eps
48        b_vec <- crossprod(X_t, Z_tilde) / sigma2_eps
49
50        if (t == 1) {
51          A <- A + diag(1 / sigma2_0 + psi[1], K)
52          b_vec <- b_vec + psi[1] * beta[, 2]
53        } else if (t == T_periods) {
54          A <- A + diag(psi[T_periods - 1], K)

```

```

55     b_vec <- b_vec + psi[T_periods - 1] * beta[, T_periods - 1]
56   } else {
57     A <- A + diag(psi[t - 1] + psi[t], K)
58     b_vec <- b_vec + psi[t - 1] * beta[, t - 1] + psi[t] * beta[, t + 1]
59   }
60
61   U <- chol(A)
62   y_vec <- forwardsolve(t(U), b_vec)
63   mu_star <- backsolve(U, y_vec)
64   beta[, t] <- mu_star + backsolve(U, rnorm(K))
65 }
66
67 for (t in seq_len(T_periods))
68   beta[, t] <- beta[, t] - mean(beta[, t])
69
70 # 3: sigma2_t
71 for (t in 2:T_periods) {
72   diff_sq <- sum((beta[, t] - beta[, t - 1])^2)
73   sigma2_t[t - 1] <- r_inv_gamma(1, a = a_sig + (K - 1) / 2,
74                                 b = b_sig + diff_sq / 2)
75 }
76
77 # 4: sigma2_eps
78 eta <- fun_eta(incontri, beta, alpha)
79 resid_sq <- sum((Z - eta)^2)
80 sigma2_eps <- r_inv_gamma(1, a = a_eps + n / 2, b = b_eps + resid_sq / 2)
81
82 # 5: alpha
83 alpha <- sample_alpha(incontri, Z, beta, sigma2_eps, sigma2_alpha,
84                      idx_t, n_t)
85
86 if (iter > burnin) {
87   save_idx <- save_idx + 1
88   beta_chain[, , save_idx] <- beta
89   sigma2_chain[, save_idx] <- sigma2_t
90   sigma2_eps_chain[save_idx] <- sigma2_eps
91   alpha_chain[, save_idx] <- alpha
92 }
93 }
94 list(beta_chain = beta_chain,
95       sigma2_chain = sigma2_chain,
96       sigma2_eps_chain = sigma2_eps_chain,
97       alpha_chain = alpha_chain)
98 }

```

CODICE A.2: Codice R per il *Gibbs sampler*, specificazione DOP.

```

1  # Gibbs sampler per la specificazione DAOP
2
3  source("funzioni.R")
4
5  # Logaritmo della densita' marginale dell'incremento sotto una
6  # componente Gamma Inversa(a, b).
7  log_dens_marginale <- function(D, a, b) {
8    a * log(b) - lgamma(a) + lgamma(a + 1 / 2) - (a + 1 / 2) * log(b + D / 2)
9  }
10
11 gibbs.multi.dinamico.ss <- function(dati, hyp, niter = 10000, burnin = 2000) {
12
13   incontri <- dati$incontri
14   K <- dati$K
15   T_periods <- dati$T_periods
16   n <- nrow(incontri)
17   Tm1 <- T_periods - 1
18
19   sigma2_0 <- hyp$sigma2_0
20   a_sp <- hyp$a_sp; b_sp <- hyp$b_sp
21   a_sl <- hyp$a_sl; b_sl <- hyp$b_sl
22   a_eps <- hyp$a_eps
23   b_eps <- hyp$b_eps
24   sigma2_alpha <- hyp$sigma2_alpha
25   c <- hyp$c
26   d <- hyp$d
27
28   beta <- matrix(0, nrow = K, ncol = T_periods)
29   sigma2_team <- matrix(0.1, nrow = K, ncol = Tm1)
30   xi <- matrix(1, nrow = K, ncol = Tm1)
31   pi0 <- rep(0.5, K)
32   sigma2_eps <- 1
33   alpha <- rep(0.3, T_periods)
34
35   idx_t <- lapply(seq_len(T_periods), function(t) which(incontri$t == t))
36   n_t <- vapply(idx_t, length, 1)
37   X_t_list <- lapply(seq_len(T_periods), function(t) fun_X_t(incontri, t, K))
38   XtX_list <- lapply(X_t_list, crossprod)
39
40   n_saved <- niter - burnin
41   beta_chain <- array(NA, dim = c(K, T_periods, n_saved))
42   sigma2_chain <- array(NA, dim = c(K, Tm1, n_saved))
43   xi_chain <- array(NA, dim = c(K, Tm1, n_saved))
44   pi0_chain <- matrix(NA, nrow = K, ncol = n_saved)
45   sigma2_eps_chain <- numeric(n_saved)
46   alpha_chain <- matrix(NA, nrow = T_periods, ncol = n_saved)
47   save_idx <- 0
48
49   for (iter in seq_len(niter)) {
50
51     # 1: Z[i]
52     Z <- sample_Z(incontri, beta, alpha, sigma2_eps)
53
54     # 2: beta[, t] a blocchi

```

```

55 psi <- 1 / sigma2_team
56 for (t in seq_len(T_periods)) {
57   idx <- idx_t[[t]]
58   X_t <- X_t_list[[t]]
59   Z_tilde <- Z[idx] - alpha[t]
60
61   A <- XtX_list[[t]] / sigma2_eps
62   b_vec <- crossprod(X_t, Z_tilde) / sigma2_eps
63
64   if (t == 1) {
65     A <- A + diag(1 / sigma2_0 + psi[, 1])
66     b_vec <- b_vec + psi[, 1] * beta[, 2]
67   } else if (t == T_periods) {
68     A <- A + diag(psi[, Tm1])
69     b_vec <- b_vec + psi[, Tm1] * beta[, T_periods - 1]
70   } else {
71     A <- A + diag(psi[, t - 1] + psi[, t])
72     b_vec <- b_vec + psi[, t - 1] * beta[, t - 1] +
73               psi[, t] * beta[, t + 1]
74   }
75
76   U <- chol(A)
77   y_vec <- forwardsolve(t(U), b_vec)
78   mu_star <- backsolve(U, y_vec)
79   beta[, t] <- mu_star + backsolve(U, rnorm(K))
80 }
81
82 for (t in seq_len(T_periods))
83   beta[, t] <- beta[, t] - mean(beta[, t])
84
85 # 3: blocco (xi[k, t], sigma2[k, t]), schema collassato
86 for (t in 2:T_periods) {
87   t <- t - 1L
88   for (k in seq_len(K)) {
89     D <- (beta[k, t] - beta[k, t - 1])^2
90
91     lw_sp <- log(pi0[k]) + log_dens_marginale(D, a_sp, b_sp)
92     lw_sl <- log(1 - pi0[k]) + log_dens_marginale(D, a_sl, b_sl)
93     p_slab <- 1 / (1 + exp(lw_sp - lw_sl))
94     xi[k, t] <- rbinom(1, 1, p_slab)
95
96     a_act <- if (xi[k, t] == 1L) a_sl else a_sp
97     b_act <- if (xi[k, t] == 1L) b_sl else b_sp
98     sigma2_team[k, t] <- r_inv_gamma(1, a = a_act + 1 / 2,
99                                   b = b_act + D / 2)
100   }
101 }
102
103 # 4: p_k
104 for (k in seq_len(K)) {
105   n_sp <- sum(xi[k, ] == 0L)
106   n_sl <- sum(xi[k, ] == 1L)
107   pi0[k] <- rbeta(1, c + n_sp, d + n_sl)
108 }
109

```

```

110   # 5: sigma2_eps
111   eta      <- fun_eta(incontri, beta, alpha)
112   resid_sq <- sum((Z - eta)^2)
113   sigma2_eps <- r_inv_gamma(1, a = a_eps + n / 2, b = b_eps + resid_sq / 2)
114
115   # 6: alpha
116   alpha <- sample_alpha(incontri, Z, beta, sigma2_eps, sigma2_alpha,
117                         idx_t, n_t)
118
119   if (iter > burnin) {
120     save_idx      <- save_idx + 1
121     beta_chain[, , save_idx] <- beta
122     sigma2_chain[, , save_idx] <- sigma2_team
123     xi_chain[, , save_idx] <- xi
124     pi0_chain[, , save_idx] <- pi0
125     sigma2_eps_chain[save_idx] <- sigma2_eps
126     alpha_chain[, , save_idx] <- alpha
127   }
128 }
129 list(beta_chain      = beta_chain,
130       sigma2_chain   = sigma2_chain,
131       xi_chain        = xi_chain,
132       pi0_chain       = pi0_chain,
133       sigma2_eps_chain = sigma2_eps_chain,
134       alpha_chain     = alpha_chain)
135 }
```

CODICE A.3: Codice R per il *Gibbs sampler*, specificazione DAOP.

Bibliografia

- AGRESTI, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: Wiley, 2nd ed.
- ALBERT, J. H. & CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.
- BAIO, G. & BLANGIARDO, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics* **37**, 253–264.
- BETANCOURT, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. arXiv:1701.02434.
- BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324–345.
- BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**, 1–3.
- CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. & RIDDELL, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* **76**, 1–32.
- CATTELAN, M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statistical Science* **27**, 412–433.
- CATTELAN, M., VARIN, C. & FIRTH, D. (2013). Dynamic Bradley–Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **62**, 135–150.
- CONSTANTINOU, A. C. & FENTON, N. E. (2012). Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *Journal of Quantitative Analysis in Sports* **8**.
- COWLES, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**, 101–111.

- DAVIDSON, R. R. (1970). On extending the Bradley–Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association* **65**, 317–328.
- DIEBOLT, J. & ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)* **56**, 363–375.
- DIXON, M. J. & COLES, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **46**, 265–280.
- DOBSON, S. & GODDARD, J. (2001). *The Economics of Football*. Cambridge University Press.
- DUANE, S., KENNEDY, A. D., PENDLETON, B. J. & ROWETH, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195**, 216–222.
- EGIDI, L., MACRÌ DEMARTINO, R. & PALASKAS, V. (2025). *footBayes: Fitting Bayesian and MLE Football Models*. R package version 2.1.0.
- EGIDI, L., PAULI, F. & TORELLI, N. (2018). Combining historical data and bookmakers’ odds in modelling football scores. *Statistical Modelling* **18**, 436–459.
- EGIDI, L. & TORELLI, N. (2021). Comparing goal-based and result-based approaches in modelling football outcomes. *Social Indicators Research* **156**, 801–813.
- EPSTEIN, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology* **8**, 985–987.
- GELFAND, A. E. & SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. & RUBIN, D. B. (2014). *Bayesian Data Analysis*. CRC Press, 3rd ed.
- GELMAN, A. & RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.
- GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

- GODDARD, J. & ASIMAKOPOULOS, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting* **23**, 51–66.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
- HIGHAM, N. J. (2002). *Accuracy and Stability of Numerical Algorithms*. Philadelphia: SIAM, 2nd ed.
- HOBBS, B. P., CARLIN, B. P., MANDREKAR, S. J. & SARGENT, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67**, 1047–1056.
- HOBBS, B. P., SARGENT, D. J. & CARLIN, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* **7**, 639–674.
- HOFFMAN, M. D. & GELMAN, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**, 1593–1623.
- HONG, H., FU, H. & CARLIN, B. P. (2018). Power and commensurate priors for synthesizing aggregate and individual patient level data in network meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **67**, 1047–1069.
- KARLIS, D. & NTZOUFRAS, I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**, 381–393.
- KARLIS, D. & NTZOUFRAS, I. (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software* **14**, 1–36.
- KARLIS, D. & NTZOUFRAS, I. (2009). Bayesian modelling of football outcomes: Using the Skellam’s distribution for the goal difference. *IMA Journal of Management Mathematics* **20**, 133–145.
- KNORR-HELD, L. (2000). Dynamic rating of sports teams. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49**, 261–276.
- KONING, R. H. (2000). Balance in competition in Dutch soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49**, 419–431.

- KOOPMAN, S. J. & LIT, R. (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **178**, 167–186.
- KOOPMAN, S. J. & LIT, R. (2019). Forecasting football match results in national league competitions using score-driven time series models. *International Journal of Forecasting* **35**, 797–809.
- KUYPERS, T. (2000). Information and efficiency: An empirical study of a fixed odds betting market. *Applied Economics* **32**, 1353–1363.
- LIU, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association* **89**, 958–966.
- MACRÌ DEMARTINO, R., EGIDI, L. & TORELLI, N. (2025). Alternative ranking measures to predict international football results. *Computational Statistics* **40**, 1899–1917.
- MACRÌ-DEMARTINO, R., EGIDI, L. & TORELLI, N. (2026). Bayesian weighted discrete-time dynamic models for association football prediction. *Journal of the Royal Statistical Society: Series C: Applied Statistics* .
- MAHER, M. J. (1982). Modelling association football scores. *Statistica Neerlandica* **36**, 109–118.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.
- MITCHELL, T. J. & BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83**, 1023–1032.
- MURRAY, T. A., HOBBS, B. P. & CARLIN, B. P. (2015). Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. *The Annals of Applied Statistics* **9**, 1549–1570.
- NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones & X.-L. Meng, eds. Chapman & Hall/CRC, pp. 113–162.

- NYBERG, H. (2014). A multinomial logit-based statistical test of association football betting market efficiency. Discussion Paper 380, HECER.
- OWEN, A. (2011). Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA Journal of Management Mathematics* **22**, 99–113.
- PARK, T. & VAN DYK, D. A. (2009). Partially collapsed Gibbs samplers: Illustrations and applications. *Journal of Computational and Graphical Statistics* **18**, 283–305.
- R CORE TEAM (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- REEP, C., POLLARD, R. & BENJAMIN, B. (1971). Skill and chance in ball games. *Journal of the Royal Statistical Society: Series A* **134**, 623–629.
- ROBERT, C. P. & CASELLA, G. (2010). *Introducing Monte Carlo Methods with R*. Springer.
- ROBERTS, G. O. & SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**, 291–317.
- RUE, H. & HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, vol. 104 of *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC.
- RUE, H. & SALVESEN, Ø. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)* **49**, 399–418.
- SHIN, H. S. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. *The Economic Journal* **103**, 1141–1153.
- SKELLAM, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society: Series A* **109**, 296.
- TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.
- VAN DYK, D. A. & PARK, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association* **103**, 790–796.

- VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. & BÜRKNER, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis* **16**, 667–718.

