

Erasmus
School of
Economics

Sports Economics: Sports Analytics

Dr. Sam Hoey
sportecon@ese.eur.nl

Erasmus University Rotterdam



Today

Learning objectives:

- 1. Develop and tune a data-driven model to predict match outcomes, evaluate its performance using out-of-sample metrics (e.g., Brier score, accuracy), and compare it against relevant benchmarks.**
- 2. Translate model predictions into a transparent betting strategy and assess its profitability and Sharpe ratio using out-of-sample testing.**
- 3. Create and share a clean replication package (code + data + README) enabling third-party replication.**

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color, located in the bottom right corner of the slide.

Today (and previous lecture)

Questions:

1. What forms of Sport Economics and Analytics are there? (*course overview*)
2. **What methods can be used to predict match outcomes?**
3. **How can we tune (hyper) parameters of our model?**
4. **How do we test the performance of our model?**

→ I will use some mentimeter quizzes throughout the lecture(s).

A handwritten signature in black ink, appearing to read "Erasmus", located in the bottom right corner of the slide.

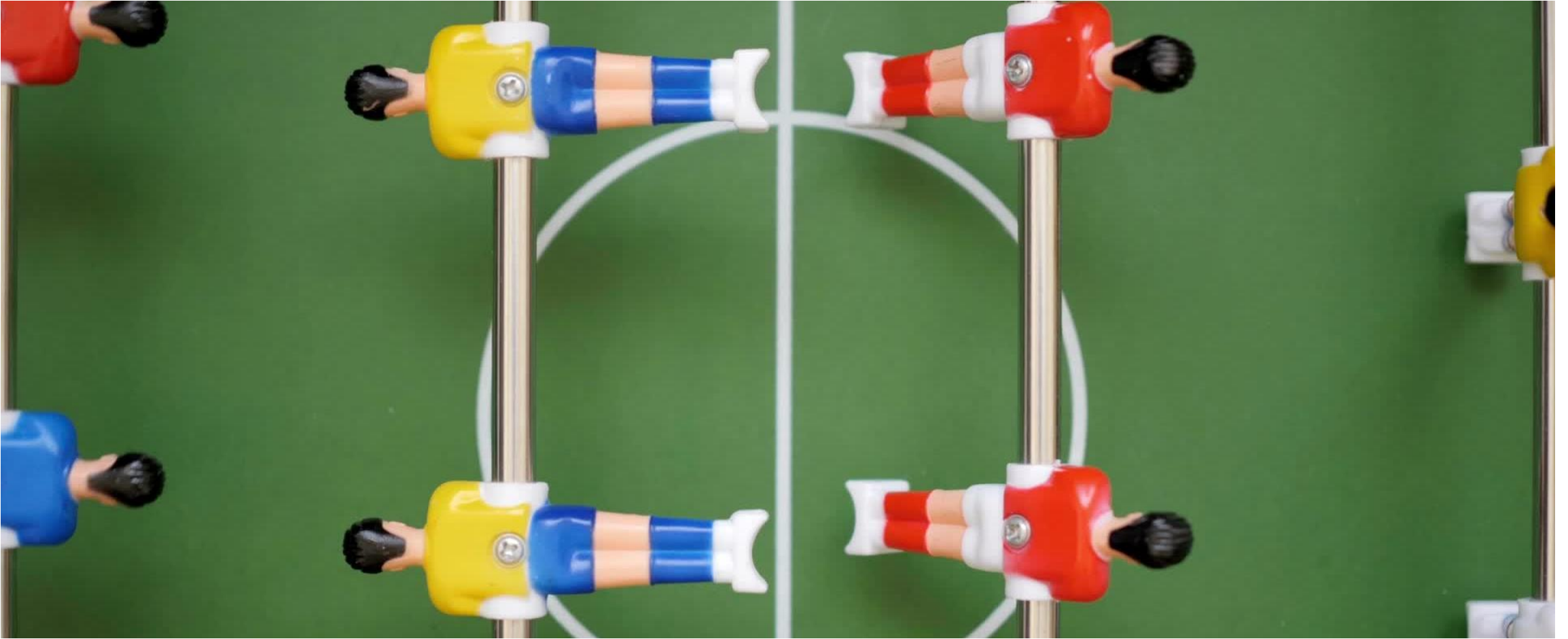
Readings

- Academic papers
 - Holmes, B., & McHale, I. G. (2024). Forecasting football match results using a player rating based model. *International Journal of Forecasting*, 40(1), 302-312.
 - Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17-29.
 - Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of forecasting*, 26(3), 460-470.
 - 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(2), 265-280.

Find them in the literature section in canvas!

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color.

Sports Analytics – Match Prediction



Ezra

What is there to predict?

Bookmakers generally provide odds for:

Match Outcome (1X2 Market)

- **1:** Home Win
- **X:** Draw
- **2:** Away Win

Proposition bets:

- Goals scored by each team
- Total goals (Over/Under)
- First goal scorer
- Correct score
- Half-time/full-time results
- Shots, corners, penalties, or player-specific events

These markets reflect expectations about team strength and can serve as a useful benchmark for model evaluation.

The screenshot shows the TOTO website interface. The top navigation bar includes links for SPORT, CASINO, LIVE CASINO, and BINGO. The main header features the TOTO logo and navigation links like SUPER WEEK, TOTO SCORE 6, TOTO EXTRA, KLANTENSERVICE, ACTIES, VERANTWOORD SPELEN, INLOGGEN, and AANMELDEN.

The main content area displays a live match page for the Netherlands Eredivisie, Feyenoord vs PSV. The score is 0-1, and the time is 46:32. The page shows various betting odds for match outcomes (1X2) and proposition bets (TOTO SPECIALS LIVE).

Match Outcome (1X2 Market) Odds:

Team	Score	Odds
Feyenoord	1	5,50
Draw	X	3,70
PSV	2	1,64

TOTO SPECIALS LIVE Odds:

Event	Odds
Ueda 1 of meer schoten op doel tussen minuut 45:00 en 54:59	5,25
Ueda 1 of meer schoten op doel tussen minuut 55:00 en 64:59	5,25
Hadj Moussa 1 of meer schoten op doel tussen minuut 45:00 en 54:59	7,00
Hadj Moussa 1 of meer schoten op doel tussen minuut 55:00 en 64:59	7,00
Sauer 1 of meer schoten op doel tussen minuut 45:00 en 54:59	9,50

The right sidebar shows a live match feed with a video player and a list of other matches to watch, including AZ vs FC Utrecht and Go Ahead Eagles vs Excelsior.

Erasmus

Match prediction methods

Generally, we use econometric/statistical methods to predict match outcomes

Three Key Stages of Match Prediction Modelling:

1. Training Stage

Choose model and estimate model parameters using historical data.

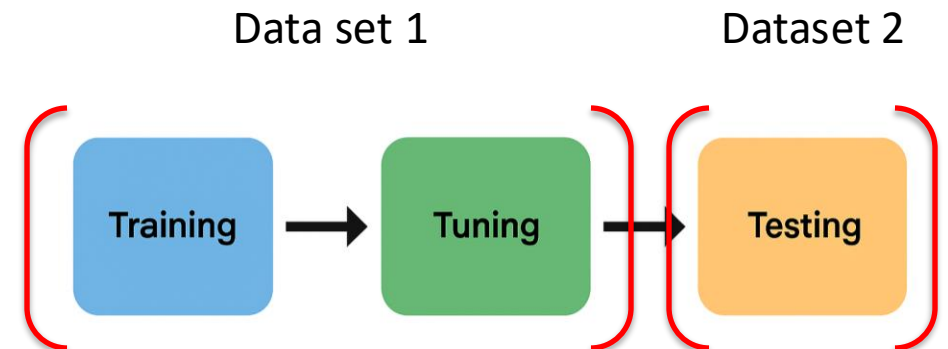
2. Tuning Stage

Adjust **hyperparameters** (e.g., window size, time decay rate, other features) to optimize predictive performance.

3. Testing Stage (next lecture)

Evaluate model performance on unseen data (test sample) to ensure robustness and avoid overfitting.

→ **Generally**, data is split into a training & tuning subset and a testing subset. **Often**, training and tuning is also split for model tuning.



Training Stage

Goal:

Build the *initial* model by estimating its parameters using **historical data**.

1. Prepare data

- Clean and structure match data (remove errors, ensure time order).
- Compute relevant features (e.g., team strength, form, ELO ratings)

2. Estimate model parameters

- Fit the chosen econometric/statistical model (e.g., OLS, ordered probit, logistic, Poisson).
- Learn relationships between covariates and match outcomes.

3. Assess in-sample fit

- Evaluate model fit (BIC, AIC, likelihood, pseudo- R^2 , residuals).
- Identify over-fitting patterns.

4. Save model outputs

- Store estimated parameters, fitted values, and diagnostics.
- Use these as inputs for tuning and validation.

Leakage: what NOT to do

- Use **end-of-season** stats to predict mid-season games.
- Build features with **future matches** (e.g., last-5 form that includes the target game).

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Summary of papers - Training

Core pipeline (all four papers):

- Build a **team strength rating** (goals model / ELO / crowd market values / player-based).
- Form match features (e.g., **rating gap** = Home – Away, plus **home** dummy).
- Convert features → **P(Home, Draw, Away)** with a **multinomial model** such as **ordered probit/logit**.

Methods used for building team strength variable:

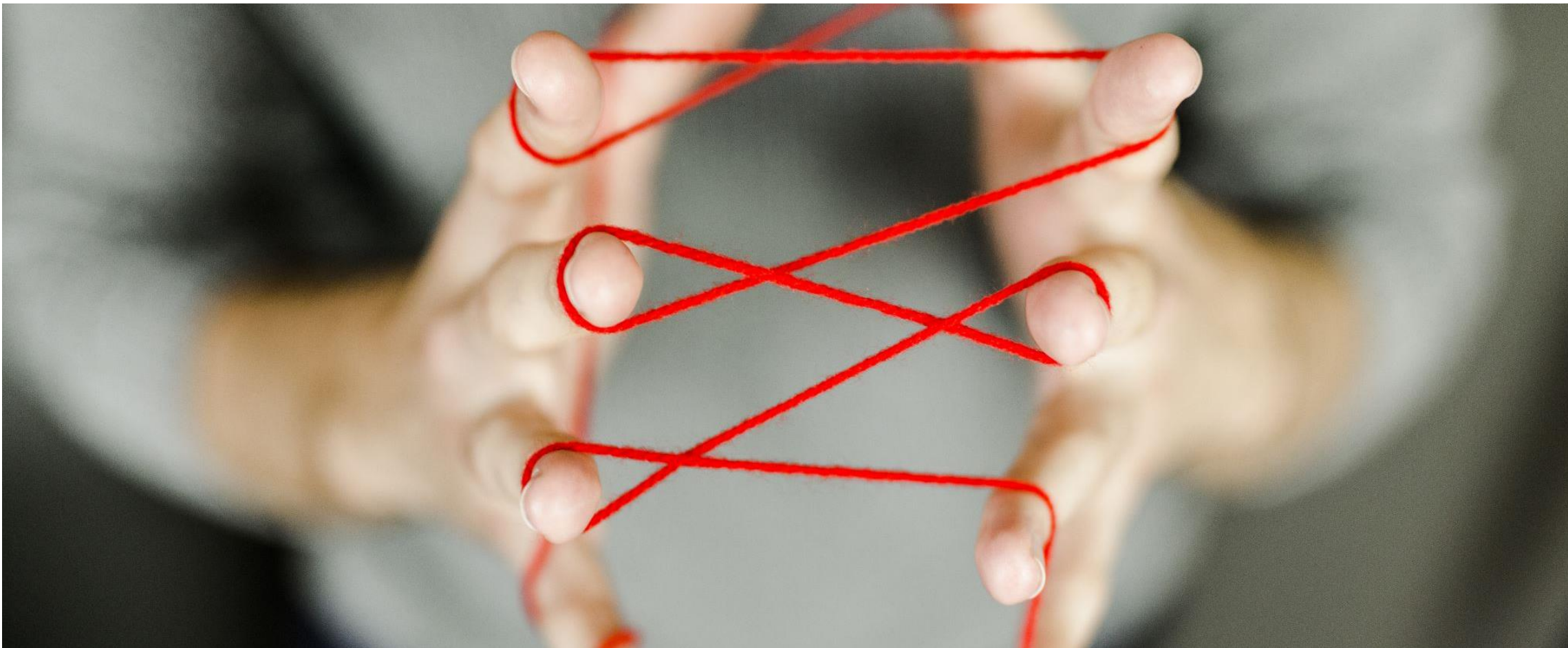
- Dixon and Coles (1997): home and away strength based on goals scored and conceded
- Artnzen & Hvattum (2010): rolling ELO ratings
- Peeters (2018): wisdom of the crowds measured by TM market values of players
- Holmes and McHale (2024): player-based model with fielded team taken as given.

Methods for within sample goodness-of-fit (Peeters, 2018):

- Psuedo R-squared
- AIC and BIC
- Log-likelihood



Tuning your model



Erasmus

Tuning and Training

Parameters vs. hyper parameters vs. features.

- Parameters are estimated during training or **learned from data**.
 - *Dixon–Coles (1997)*: team **attack, defence and home advantage**,.
 - *Arntzen & Hvattum (2010)*: **ordered-logit β** , relating elo difference to outcomes, and **cutpoints τ_1 and τ_2** .
 - Can be updated in test data set using **fixed update rule (we do not do this in assignment)**.
- Hyper parameters are **chosen** by the researcher
 - **Window length**: how many past matches enter estimation.
 - **Number of seasons included for player ratings** (Holmes & McHale, 2024).
 - **Recency/decay**: weight on past games (Dixon & Coles, 1997; Holmes & McHale, 2024).
 - **ELO update size**: learning rate $k \rightarrow$ how much ratings move after each game (Arntzen & Hvattum , 2010).
 - Not updated in test data set generally.
- Features/variables are variables that feed into models
 - **Examples**: ELO score, player ratings, Transfermarkt values
 - Can be updated in the test data set using **fixed update rule**

→ We *estimate* parameters and *tune* hyperparameters

→ To tune hyper parameters, we use *benchmarks*.



Benchmarks (needed for tuning)

Accuracy or Success ratio (*higher is better*)

- Evaluates how often the model predicts the right outcome by looking at highest probability outcome and comparing to actual outcome.
- Calculated as the percentage of cases that the model predicts right.
- **Pros:** Simple, intuitive headline metric.
- **Cons:** Ignores probability quality (51% = 99%); sensitive to outcome (im)balance.

Brier score (Quadratic Loss, MSE, or Square of RMSE, *lower is better*)

- Evaluates Quality of **probability forecasts** — penalizes being over/under-confident relative to what actually happened.
- **How it's computed (multi-class 1X2):**

$$Brier = \frac{1}{N} \sum \sum_{k \in \{H,D,A\}}^N (p_{ik} - o_{ik})^2$$

Where p_{ik} = predicted prob for outcome k (*home, draw, away*), $o_{ik} = 1$ if k occurred, else 0.

- Some papers divide by $K = 3$. **Do not do this in report (we want results to be comparable)**
- **Interpretation:** 0.67 \rightarrow model as good as random guess, 0 \rightarrow model predicts perfectly, good score is context dependent
- **Why use it?**
 - Uses full probability distribution (not just top pick)
 - Rewards good well-tuned models



Examples – Tuning decay Parameter

Dixon and Coles (1997) and Holmes and McHale (2024) **tune their time decay parameter.**

- **Dixon and Coles (1997)** less weight on observations further in the past in likelihood function
- **Holmes and McHale (2024)** less weight on player ratings/games (AWS) further in past when calculating player ratings
- Formula: $w_{it} = \exp(-\phi t)$, where t is number of days prior to calculation day.

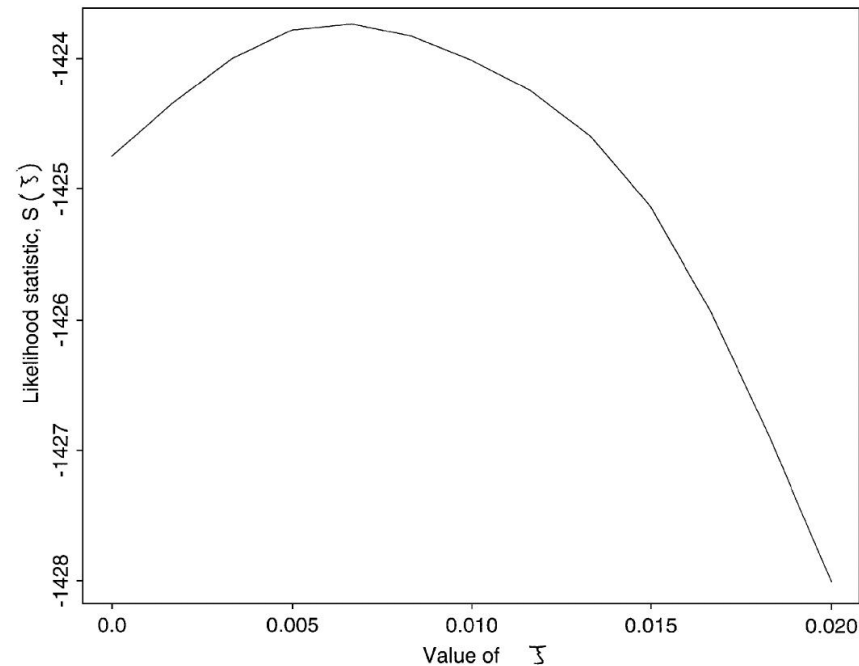
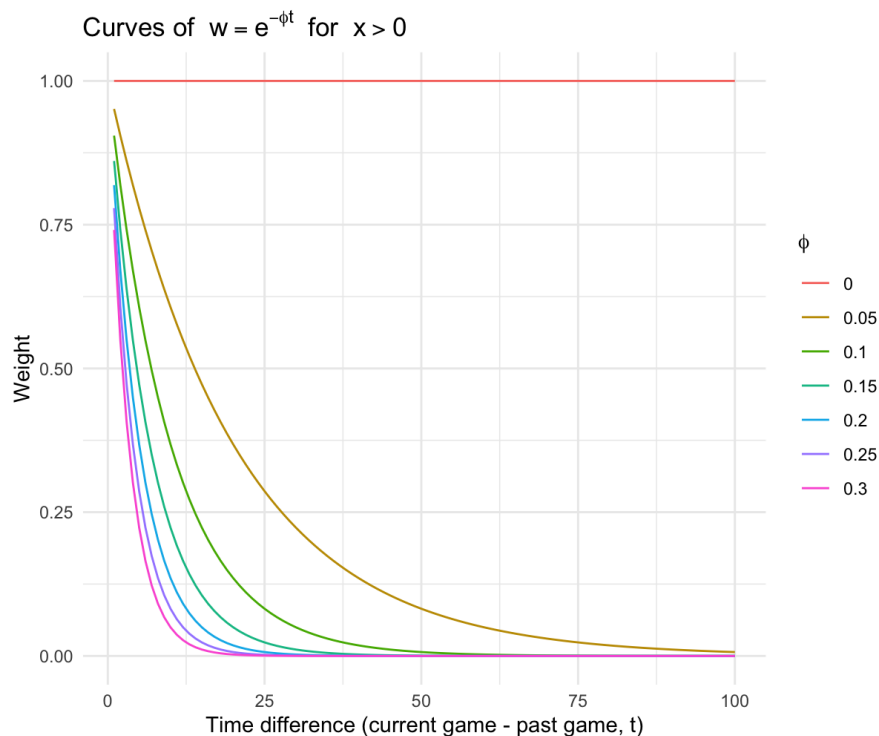


Fig. 1. $S(\xi)$ versus ξ : the maximum occurs at $\xi = 0.0065$

Dixon and Coles (1997)

Erasmus

Examples – Tuning updating Parameters

Arntzen and Hvattum (2010) tune two parameters at once

Elo formula:

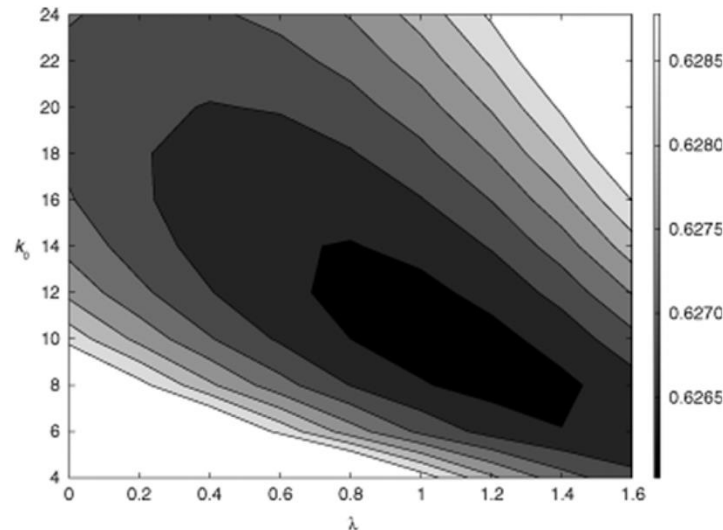
$$\ell_1^H = \ell_0^H + k(\alpha^H - \gamma^H).$$

k governs the amount of updating;
high k = a lot of weight on most recent game;
low k = less weight on most recent game

Goal adjusted k-factor:

$$k = k_0(1 + \delta)^\lambda,$$

k_0 and λ govern updating taking into account goal difference δ



They end up using:

$$k_0 = 10$$

$$\lambda = 1$$

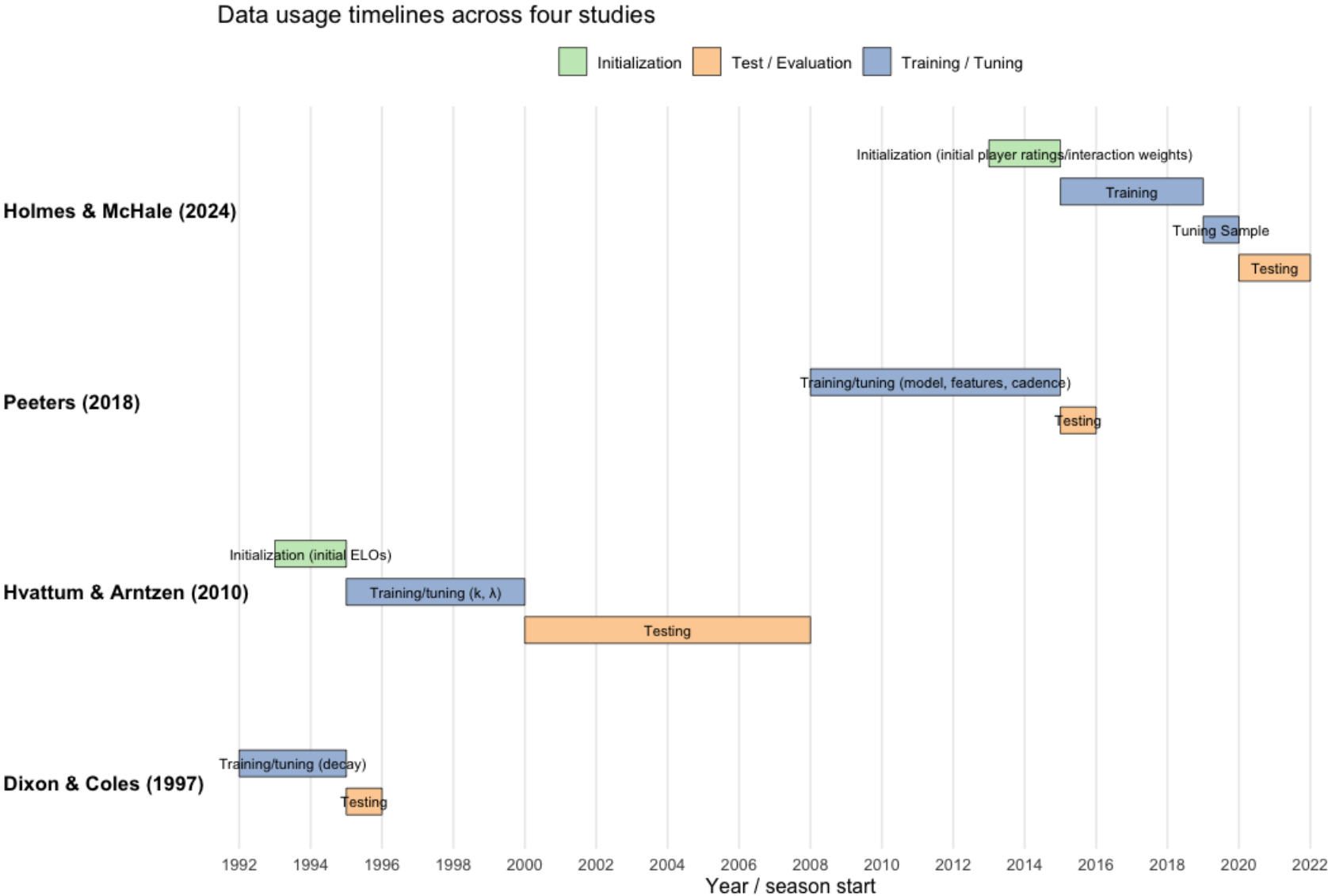
Arntzen

Fig. 1. Observed average quadratic loss when calibrating the two free parameters of the **ELO_g** method, k_0 and λ .

Model Testing

Erasmus

Training and Tuning Stage Summary



zafun

Model Testing

- **What testing is:**

Check how well your frozen model (parameters + hyperparameters) predicts unseen games → no refitting, no retuning.

- **What overfitting is:**

When a model learns noise/idiosyncrasies specific too the training data → within sample metrics look great, but out of sample predictions are poor.

- **Why it happens (match prediction):**

- Too many features based on few observations → noisy parameters
- Overly aggressive time decay/k (model only considers last few games) .

- **How testing guards against it:**

A strict chronological test sample (e.g., 2021–2023) reveals whether patterns learned on earlier data (e.g. 2011–2020) actually generalize.

- **Tell-tale signs:**

Big gap: Train/Validation >> Test on Brier Score; overconfidence (probabilities too close to 0/1)

- **Good practice to avoid it:**

- **Train and tune** parameters and hyperparameters on training & tuning block.
- **Freeze** parameters & hyperparameters before testing.
- **Keep it simple:** start with few, well-motivated features.



Keeping it simple (also relevant for training)

Don't: Add a dummy like “**vs_Bayern**” for each opponent.

Do instead: Use a **strength rating** for the opponent (e.g., ELO or rating gap).

Why: Facing a particular opponent is rare → the model memorizes, not learns.

Don't: Add home season-specific home advantage parameter for each team.

Do instead: Use one overall home-advantage indicator.

Why: Granular season interacted dummies overfit as they are estimated on few games.

Don't: Indicators for whether player X is injured

Do instead: Team strength ratings based on players in lineup

Why: for some players injuries might be rare → model memorizes what happened in those specific games.

Morale of the story: Prefer pooled/continuous features over rare, specific dummies so the model generalizes and avoids overfitting.

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Join at menti.com | use code 5605 4055

Mentimeter

Your connection seems to be unstable. Some features may not work as expected.

Which feature is **most likely to cause overfitting** in a match-outcome model?

0

Opponent rating gap (home – away)

0

Overall home-advantage indicator

0

Referee strictness index

0

ELO with very high k factor

0 / 31



SH



Menti

New presentation



What can we use to predict match outcomes?

past results
toto xg
injuries
home advantage
form
opponent rating gap
overall home advantage indicator
referee strictness index
elo with very high k factor

Which feature is **most likely to cause overfitting** in a match-outcome model?

0 0 0 0
Opponent rating gap (home – away) Overall home advantage indicator Referee strictness index Option X

One concept still fuzzy + one you can explain to a friend.

Erasmus

Evaluating & Benchmarking on Test Data

Metrics (same as tuning): report Brier (primary) and Accuracy on the frozen model.

Benchmarks to compare against:

- Uniform random: $p(\text{home, draw, away}) = (1/3, 1/3, 1/3) \rightarrow \text{Brier} = 0.667$.
- Bookmakers' implied probabilities (more later).
- Simple baselines: majority class (Accuracy), historical frequencies (home win, draw, away win).

Model A vs B: Paired t-test on Brier

- For each game t , compute a metric for both methods (e.g., Brier): b_t^A vs b_t^B .
- Run a paired t-test on the per game differences to evaluate whether they are statistically different.
- Equivalently: let $d_t = b_t^A - b_t^B \rightarrow \text{test } H_0 = E[d] = 0$

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Arntzen & Hvattum (2010) – Testing Sample Metrics

Table 5

Average values (AVG) and standard deviations (STD) for informational (L^I) and quadratic (L^2) loss on 15,181 matches from the English league system, obtained using eight different prediction methods. The p -values from matched pair t -tests are also reported, comparing each method with ELO_g .

	L^I			L^2		
	AVG	STD	p -value	AVG	STD	p -value
UNI	1.5850	0.0000	0.0000	0.6667	0.0000	0.0000
FRQ	1.5443	0.3791	0.0000	0.6469	0.1870	0.0000
GOD _b	1.5135	0.4824	0.0000	0.6321	0.2371	0.0000
GOD _g	1.5129	0.5136	0.0000	0.6315	0.2472	0.0000
ELO _b	1.5018	0.5030	0.0001	0.6266	0.2475	0.0001
ELO _g	1.4995	0.5060	NA	0.6256	0.2489	NA
AVG	1.4917	0.4772	0.0000	0.6219	0.2340	0.0000
MAX	1.4910	0.4955	0.0000	0.6217	0.2415	0.0000

Brier Score comparison against preferred model ELO_g

Model Testing with Betting Odds

Erasmus

Evaluating & Benchmarking on Test Data

What are odds? (decimal format)

- Decimal odds o : your **total return per \$1 staked** if the outcome wins.
Example: $o = 2.40 \rightarrow$ win returns \$2.40 (profit \$1.40).

Implied probability

- $p_k^{book} = \frac{1}{odds_k}$ for outcome k (home/draw/away)
- For a fair market, $p_h^{book} + p_d^{book} + p_a^{book} = 1$

Bookmaker margin (overround)

- $R = \sum_k p_k^{book} = p_h^{book} + p_d^{book} + p_a^{book}$ (typically $R > 1$)
- $(R - 1)$ is the margin spread across outcomes (or the vig)

Converting odds (removing margin)

$$p_k^{fair} = \frac{p_k^{book}}{R}$$

Example:

Odds: Home **2.00**, Draw **3.50**, Away **3.70**

Implied probabilities: 0.500 (home) , 0.286 (draw) , 0.270 (away)

Overround: $R = 0.500 + 0.286 + 0.270 = 1.056$ ($\approx 5.6\%$ margin)

Fair probabilities:

$$p_{home}^{fair} = 0.500/1.056 \approx 0.474$$

$$p_{draw}^{fair} = 0.286/1.056 \approx 0.270$$

$$p_{away}^{fair} = 0.270/1.056 \approx 0.256$$

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Expected value (EV) betting

What is EV?

- With decimal odds o and your model probability p :
- $EV = p \cdot o - 1$ is the expected profit per \$1 staked.
 - $EV > 0 \Rightarrow$ positive edge (profitable in expectation)
 - $EV = 0 \Rightarrow$ break-even
 - $EV < 0 \Rightarrow$ negative edge
- Crucial: EV hinges on your p being the true probability

Selection rule (typical):

- Bet outcome k only if $EV_k \geq \tau$ (in principle no limits to τ , but typical is from 0% to 5%)
- You can even tune τ in your training/tuning sample to max profits.

Staking (in academic papers)

- Bet **unit stake/flat stake** on most likely outcome when EV threshold is met (default in many papers)
 - Max one bet per match
- Kelly criterion
 - Let $b = o - 1$ (o = bookmaker odds), p = model probability and $q = 1 - p$
 - Kelly fraction: $f^* = \frac{bp - q}{b}$
 - Modified application (Holmes & McHale, 2024): tune EV threshold \rightarrow if met stake 1 unit $\times f^*$ (no bankroll)
 - Can place multiple bets per match.



Performance measures

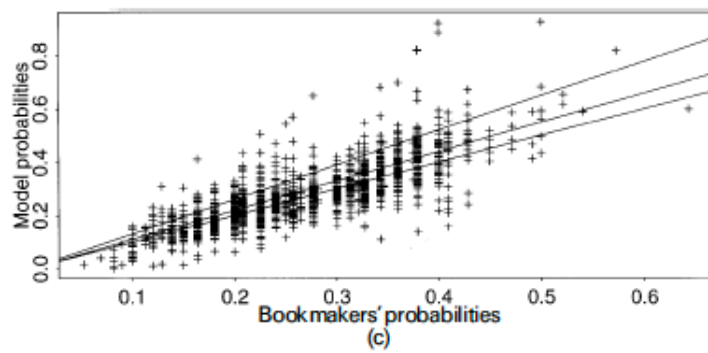
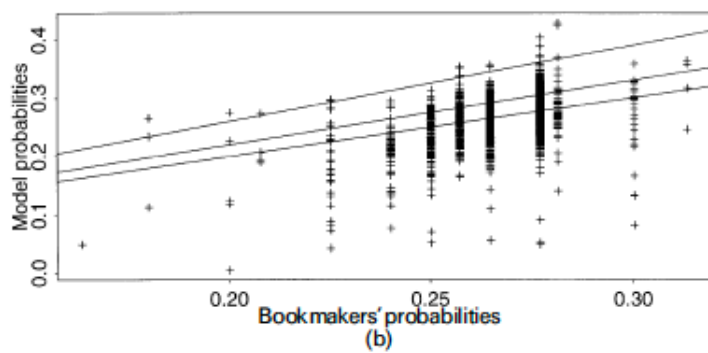
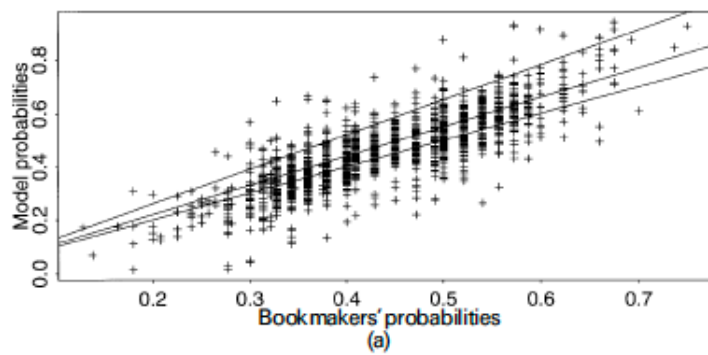
Typical reporting (see Holmes and McHale, 2024):

1. Number of bets placed under betting rules (N)
2. EV threshold used
3. Accuracy (hit-rate): how often did bet win (winning bets/N)
4. Profits (returns – stakes)
5. ROI (overall): *return on investment* $= \frac{\text{profits}}{\text{stakes}} \times 100\%$
6. Sharpe $= \frac{ROI(overall)}{sd(per\ bet\ ROI)} * \sqrt{N}$ (Risk adjusted performance → above 1 is good)

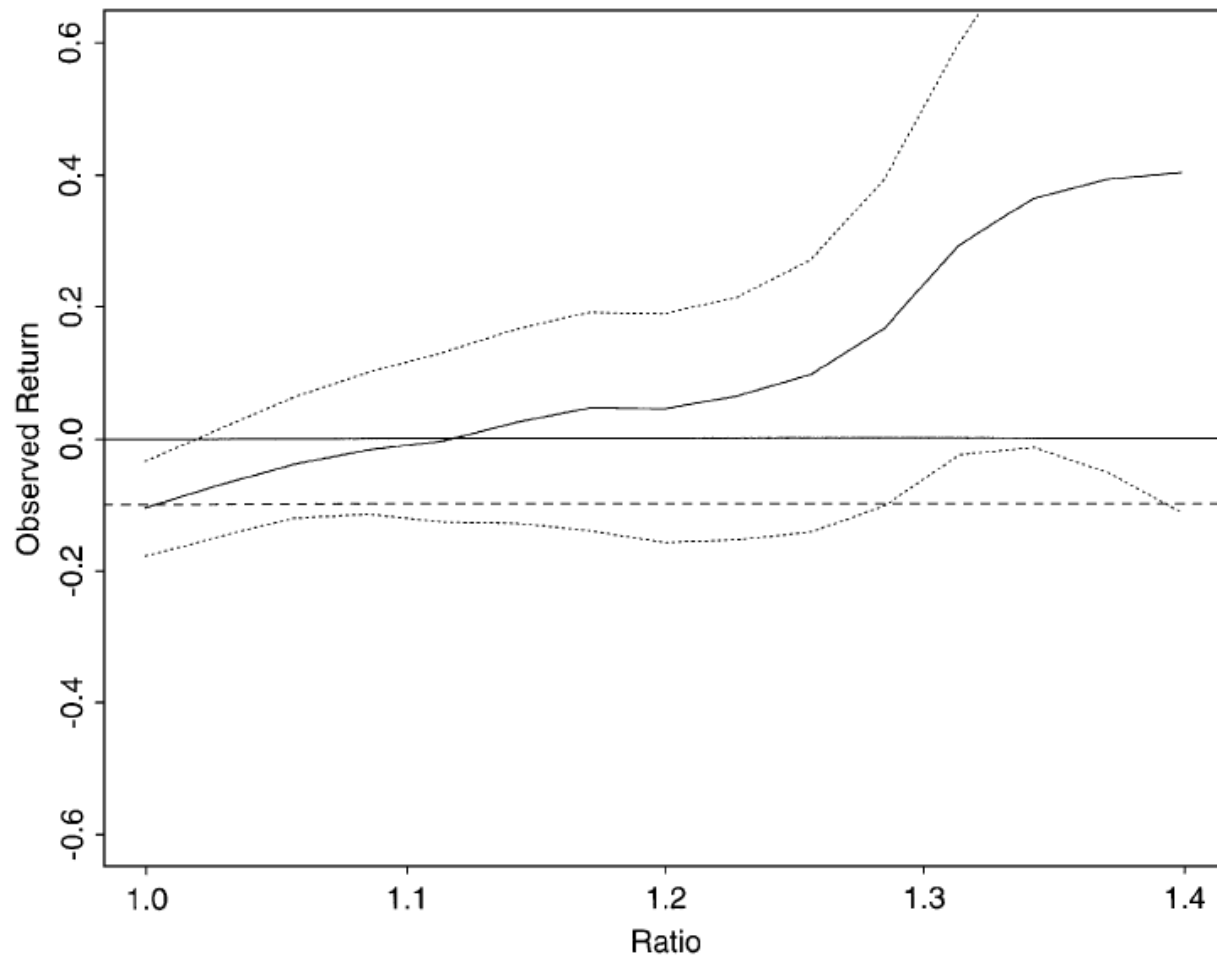
A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Dixon and Coles (1997)

Model Predictions vs Bookies



Expected RIO for different EV Thresholds, $(1 + \tau)$ here



Why do the confidence bounds expand?

Holmes and McHale (2024)

Table 6

Results for several betting strategies using the skellam_{full} model.

Strategy	t	N	Accuracy (%)	Stakes	Profits	ROI (%)	Sharpe
Kelly	0.1866	556	24.10	65.36	7.81	11.96	1.07
Kelly	0.0000	1457	29.44	105.42	6.03	5.72	1.02
Flat	0.1760	199	37.19	199.00	9.04	4.54	0.45
Flat	0.0000	568	45.25	568.00	16.93	2.98	0.59
Flat		1350	52.00	1350.00	-32.56	-2.41	-0.87

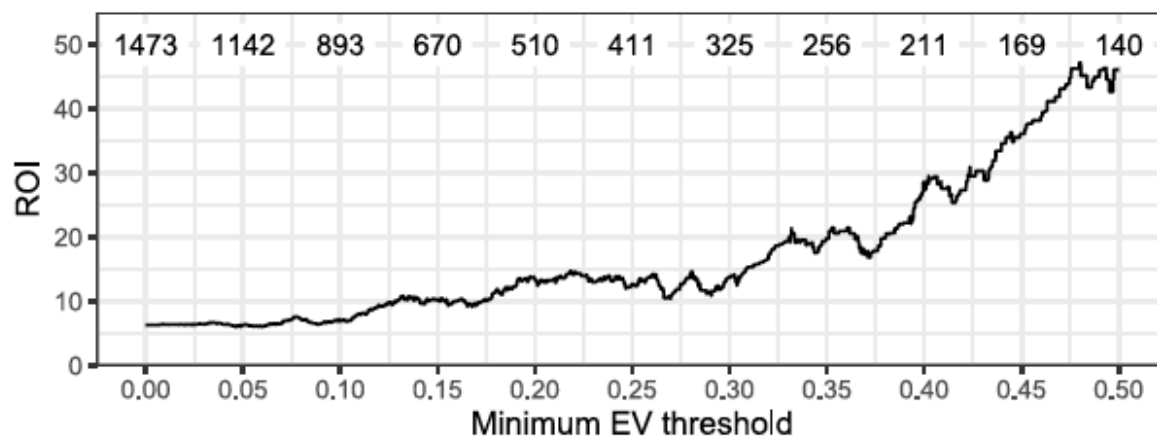


Fig. 5. Plot displaying the ROI that would be achieved using the skellam_{full} model for betting under the modified Kelly strategy for different minimum expected value thresholds.

Erasmus

Join at menti.com | use code 5605 4055

Mentimeter

What is true about betting?

0

Setting a higher EV threshold leads to higher profits.

0

Higher accuracy (hit-rate) means higher profits.

0

Sharpe ratio quantifies risk adjusted returns.

0 / 31



Menti

New presentation

What is true about staking?

0 0 0

Setting a higher EV threshold leads to higher profits. Higher accuracy (hit-rate) means higher profits. Sharpe ratio quantifies risk adjusted returns.

Which feature is most likely to cause overfitting in a match-outcome model?

0 0 0 0

Opponent rating gap (home - away) Overall home advantage indicator Reference strictness index Option 4

Erasmus

Training, Tuning and Testing Paper Summary

Study	Model Type	Training (data & method)	Validation (tuning)	Test (holdout)
Dixon & Coles (1997)	Bivariate Poisson for home/away goals	1992–1995 English league; estimate team attack/defence + home advantage ; exponential time weighting; derive 1X2 from goal model	Within 1992–1995 : choose decay parameter (and low-score tweak) by maximizing predictive score on past matches (time-ordered)	1995–1996 season; out-of-sample probabilities vs bookmaker odds
Hvattum & Arntzen (2010)	ELO ratings + ordered logit (rating gap → 1X2)	Initialize ELO on early seasons 1993/94–1994/95 ; estimate ordered-logit using ELO on 1995/96–1999/00	Tune k_0 (learning rate) and λ (margin amplifier) on pre-2000 seasons using Quadratic Loss (brier)	2000/01–2007/08 ; evaluate predictive scores and betting simulations
Peeters (2018)	Ordered probit/logit using Transfermarkt squad values (crowd signal)	International matches (UEFA/CONMEBOL) 2008–2014 ; predictors: pre-match log squad value, home, squad size; rolling estimation	Peeters (2018) does limited tuning: comparing link (probit vs logit), feature sets, and daily vs monthly updating	Out-of-sample windows (held-out periods, e.g., later seasons/tournaments); compare to bookmakers, ELO, FIFA
Holmes & McHale (2024)	Multinomial with player-based team strength (incl. pair/position effects)	Pre-train positional/interaction models 2013/14–2014/15 ; training/tuning window 2015/16–2020/21 using player ratings	Inside 2015/16–2020/21 : first 80% train , last 20% of that block validate ; tune decay parameter , shrinkage, interaction set, EV threshold	2020/21–2021/22 ; fixed parameters; report Brier/accuracy and betting vs bookmakers

Assignments

Erasmus

Assessment – Group assignment (40% of final grade)

Report 1 — Match-Outcome Prediction (30% of final grade)

- **Deadline:** Mon 24 Nov, 12:00 (noon)
- **Data:** NHL data provided
- **Tasks:**
 - Build a probability based model to predict NHL match outcomes.
 - Provide model evaluation scores and compare against relevant benchmarks.
 - Provide results for different betting strategies based on probability model.
- **Deliverables:** PDF report, code, log files and dataset.

Report 2 — Peer Evaluation and Verification (10% of final grade)

- **Deadline:** Fri 5 Dec, 12:00 (noon)
- **Data:** After checks I will distribute replication packages from assignment 1
- **Tasks:**
 - **Replicate another group's results** from their code and **verify their benchmarks**.
 - Compare **your model vs other group**; explain similarities/differences.
 - Document any replication issues and propose fixes.
- **Deliverables:** PDF report & replication log.

Hackathon:

In lecture 10 we will compare results across groups and highlight what strategies/prediction models worked best.

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color.

Assignment 1 – Overview (See Assignment for more detail)

1. Predict **Home / Tie / Away** for using National Hockey League data provided on canvas.
2. Show your **decision process**: alternatives considered → final choice.
3. Evaluate with **Brier, Accuracy** (out-of-sample).
4. Convert predictions into a **transparent betting strategy**; report **profitability & Sharpe**.
5. **Groups**: 4–5 students · **Limit**: 11 pages (1.5 spacing, font 11) · **Deadline**: Mon 24 Nov, 12:00.
6. Include a short **AI reflection** (how you used ChatGPT, what you accepted/rejected).

A stylized, handwritten-style logo for Erasmus, featuring a large, flowing 'E' followed by the word 'Erasmus' in a cursive script.

Assignment 1 - Data Split & Workflow (See Assignment for More detail)

1. Train + Validate: 2011–2020

- Try models; **tune**: window length, **decay**, feature set/variables, thresholds.

2. Test (strict holdout): 2021–2023

- **No** tuning/refitting after freeze → parameters and hyperparameters fixed.

3. Betting simulations: **only** on 2021–2023 **after** tuning is frozen.

4. **Report**: training & tuning (for model and hyperparameter choices) and **final test** results (for performance).

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color, located in the bottom right corner of the slide.

Assignment 1 – Structure (see assignment for more detail)

Introduction

- Problem & contribution; modeling options considered (3+ papers cited)
- Final model preview & why (brief)

Data

- Sample: NHL games, **Train/Validate 2011–2020, Test 2021–2023**
- Sources, data cleaning, key variables, summary tables/figures

Methodology

- Model(s) and features, **tuning** (e.g. window width, decay)
- Odds → rescale probabilities for bookmaker benchmark
- Anti-leakage design; reproducibility choices

Model Performance (Test: 2021–2023)

- **Brier & Accuracy** vs. benchmarks (incl. bookmakers)
- Betting simulation: rules (EV threshold, staking), **ROI/Sharpe**

Conclusion

- Main findings; limitations

AI Reflection (last page)

- How AI helped; accepted vs. rejected suggestions; verification steps



Assignment 1 – Rules of the Game (see assignment for more detail)

1. **Fixed split** into training & tuning data and test data (2021 season onwards)
2. **Freeze** all parameters & hyperparameters **before** touching the test set.
3. **Only pre-game info** at prediction time (no end-of-season stats for earlier games).
4. **Time-updating features/variables allowed** in test data (e.g., rolling form/ratings) if updates follow the **pre-specified rule**.
5. **Player-based models:** you may treat the **listed lineup** as known pre-puck drop.
6. Keep full **reproducibility**: fixed seeds if needed , well documented, one-click run after Working Directory change.
7. AI is allowed (and encouraged), but use of it must be documented.

A handwritten signature in black ink, appearing to read 'Erasmus', is located in the bottom right corner of the slide.

Assignment 1 – Deliverables (see assignment for more detail)

Report (PDF): se_assignment1_<group_number>_report.pdf

Code:

- **R:** se_assignment1_<group_number>_code.R
- **Stata:** se_assignment1_<group_number>_code.do

README (PDF): se_assignment1_<group_number>_README.pdf (how to run, software, seeds)

Data (CSV): se_assignment1_<group_number>_data.csv

- Make it **end-to-end runnable** without manual steps; document everything clearly.

A stylized, handwritten-style logo of the word "Erasmus" in a dark blue or black color, located in the bottom right corner of the slide.

Assignment 1 – Evaluation Criteria (see assignment for more detail)

- 1. Model choice process:** alternatives tested, **tuning strategy**, clear reasoning.
- 2. Benchmarking & metrics:** **Briers, Accuracy**; compare to **relevant benchmarks**.
- 3. Betting strategy:** correct implementation; transparent rules (**EV threshold**, staking, cap); report **ROI & Sharpe** (test set only).
- 4. Clarity & documentation:** well-written, cited relevant papers; neat layout.
- 5. Reproducibility:** code/data/README produce identical results.

A handwritten signature in black ink, appearing to read "Erasmus", is located in the bottom right corner of the slide.

Assignment 2 – Overview (see assignment for more detail)

Role: Independent verifier of another group's model and replication package.

Verification (2021–2023 test only):

- Run their code **as provided** (no edits/tuning).
- Reproduce **Brier** and **Accuracy**; record exact values.
- If non-replicable, **document discrepancies** (seeds, versions, missing files, paths).

Evaluation (method & practice):

- Is the **README** clear? Does code run **end-to-end**?
- Does the design follow good practice (**chronological split**, **freeze before test**, **no leakage**, sensible **window/decay**, odds handling)?
- Are results **robust** and **interpretable**? Any red flags?

Reflection:

- What would you **adopt** or **change** in your own workflow? Why?

Deliverables (3 pages max):

- **Report PDF:** verification results, evaluation, reflection.
- **Run log:** proof of execution / where it failed.
- **Practicalities:** same group as A1; 1.5 spacing, font 11; **deadline: Fri Dec 5, 12:00**; include brief **AI reflection** if used.

Note: this report has no influence on Assignment 1 grading.



Assignment 2 – Evaluation Criteria (see assignment for more detail)

1. Verification of results:

- Successfully reproduce out-of-sample Brier & Accuracy on 2021–2023, or give a clear, evidenced account of discrepancies.

2. Methods & reproducibility:

- Evaluation of model design & tuning (leakage control, freeze, window/decay, odds handling).
- Evaluate reproducibility: README clarity, one-click code run.

3. Reflection & adaptation:

- Concrete, actionable improvements you'll make to your own model/workflow or to theirs.

4. Clarity & documentation (presentation):

- Tight, well-structured ≤ 3 pages; labeled figures/tables; precise reporting of metrics and any errors/logs.

Musts: do not modify their model and attach a run log.

A handwritten signature in black ink, appearing to read "Erasmus", is located in the bottom right corner of the slide.

Data assignment clinics

- Voluntary opportunity to get feedback for your group
 - Private Q&A with 1 teacher
 - Only assignment questions
- Planning:
 - Data clinic 1: Wednesday 12 November (scheduled appointments on Teams/Campus)
 - Proposal deadline: Monday 27 November
 - Data clinic 2: Wednesday 19 November (scheduled appointments on Teams/Campus)
 - **Report 1 deadline: 24 November 12h00**
 - Data clinic 3: Wednesday 3 December (scheduled appointments on Teams/Campus)
 - **Report 2 deadline: 5 December 12h00**

!!!! Send e-mail to sportecon@ese.eur.nl at least 2 days before to set appointment for your group!!!!

The Erasmus logo, featuring the word "Erasmus" in a stylized, cursive script.

Join at menti.com | use code 5605 4055

Mentimeter



SH



One concept still fuzzy + one you can explain to a friend.

All responses to your question will be shown here

Each response can be up to 200 characters long

Turn on voting to let participants vote for their favorites

Menti

New presentation



One concept still fuzzy + one you can explain to a friend.



0 / 31



Erasmus