Sport Economics Assignment 1

Sam Hoey

October 2025

1 Learning Objectives

1. Develop and tune a data-driven model to predict NHL game outcomes, evaluate its performance using out-of-sample metrics (e.g., Brier score, log loss, 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 and in turn replicate another researcher's work.

2 Practicalities

1. Work in a group of 4-5 students
2. All Tables and Figures must be clearly labeled, references and notes added where necessary.
3. Use of ChatGPT is permitted (and encouraged) provided you write a reflection on how it helped you in your assignment (writing, coding, ideas generated). It goes without saying, check it's suggestions. Copying chatgpt suggestions without clear documentation and demonstrated understanding will weigh heavy in the grading.
4. Limits: 11 pages all inclusive, 1.5 line spacing, 11 font size. The last page is reserved for the "use of AI" reflection and cannot be replaced by other content if you do not use AI.
5. Deadline: Monday 24 November noon

3 Assignment

3.1 Content

In this report you will build a model that produces *game-level* predictions for NHL matches *at the end of regulation time*. Your model must assign probabilities to the three possible results after 60 minutes: **home win**, **tie**, and **away win**.[1]

In writing the report, it is important to show your decision-making process. Outline the different models you considered and explain why you ultimately chose one model. Use appropriate evaluation criteria (e.g., Brier score, log loss, accuracy score) to support your choice. Clearly separate the data: **train/validate on 2011–2020** (use this period to try alternatives and **tune** settings such as window length, decay, and feature set), and keep **2021–2023** as a strict **test** period that is not used for tuning or refitting.

Finally, evaluate your model on the **test period (2021–2023)** specified above. Run any betting strategy simulations *only* on this test set—after all tuning is completed on 2011–2020—and clearly state your rules (e.g., bet only when expected value is positive, flat stakes, or fractional Kelly with a cap).

Peer verification in Assignment 2. In the follow-up assignment, a peer group will review your replication package and independently verify your *out-of-sample* results. They will run your code on the test period, reproduce your reported metrics (Brier & Accuracy scores), and flag any discrepancies. To facilitate this, your submission must be fully reproducible: include a clear README with run instructions, fixed seeds if needed, software/packages used, and scripts that execute end-to-end without manual intervention.

Hackathon. You must include both an out-of-sample Brier score and a standard accuracy metric in your report.[2] In the final lecture, we will compare these results across groups to evaluate which models performed best and which features contributed most to their success. Only groups with valid, reproducible models will be included in this comparison.

3.2 How to structure the report

The structure should broadly follow:

Holmes, B., & McHale, I. G. (2024). Forecasting football match results using a player rating based model. *International Journal of Forecasting*, *40*(1), 302–312.

[1] Overtime and shootouts are ignored here because bookmakers also quote odds for regulation-only outcomes.

[2] Accuracy is defined as Accuracy $= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{y}i = y_i)$, where $\hat{y}i$ is the predicted class and y_i is the actual outcome. The multi-class Brier score is defined as Brier $= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(p_{ik} - o_{ik})^2$, where p_{ik} is the predicted probability for class k and o_{ik} equals 1 if outcome k occurred and 0 otherwise.

and include:

1. **Introduction**: explain your project and briefly overview the methods you considered. Include at least three papers that informed your approach. (See Sections 1–2 of Holmes & McHale (2024))

2. **Data**: describe the dataset and provide descriptive statistics for key variables. (See Section 3 of Holmes & McHale (2024))

3. **Methodology**: explain how you tuned and selected the final model (e.g., window sizes, decay, feature set) and the odds-to-probabilities method used for the bookmaker comparison. (See Sections 4–6 of Holmes & McHale (2024))

4. **Model performance**: report predictive accuracy and compare your chosen model to alternatives and relevant benchmarks; include test-period results (2021–2023) and betting simulation outcomes. **Crucial:** include out-of-sample brier score and accuracy statistic. (See Section 7 of Holmes & McHale (2024))

5. **Conclusion**: summarize your findings and limitations. (See Section 8) of Holmes & McHale (2024)

6. **AI reflection**: reflect on your use of AI, how it helped you, what suggestions you accepted, which you rejected, etc.

3.3 Rules of the game

1. **Time split.** Use a chronological split: **train/validate on 2011–2020** and **test on 2021–2023**. Do not shuffle data.

2. **Freeze before testing.** All *model parameters* (e.g., ordered-probit coefficients and cutpoints) and chosen *hyperparameters* (e.g., rolling window width, decay/half-life, rating update settings) must be fixed before evaluating out-of-sample performance on 2021–2023.

3. **Time-updating features allowed.** Input variables (e.g., rolling form, ratings) may update during the test period using only information available up to each game, and according to an update rule fixed during training/validation.

4. **No leakage.** At prediction time you may use only information that would have been known before the game. For example, do not use end-of-season standings when predicting games played earlier that season.

5. **Player-based models.** If you use player-level features, you may treat the listed lineup for that game as known prior to puck drop (the sole exception to the no-leakage rule).

3.4 Deliverables

1. Report (PDF) with: model description, in-sample diagnostics, clear description of the decision-making process, and methods used.
   Filename: `se_assignment1_<group_number>_report.pdf`

2. Code used to generate outputs (Stata do-file or R script).
   Filenames:

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

3. README file: file describing all files except your report in your replication package and their use.
   Filename: `se_assignment1_<group_number>_README.pdf`

4. Dataset needed to run your code. Must be in csv format.
   Filename: `se_assignment1_<group_number>_data.csv`

3.5 Evaluation criteria

This assignment is worth 30% of your final grade for sporteconomics. You get points for the following parts:

1. **Model choice process:** use of tuning process of some sort, different models considered, clear documentation and arguments.

2. **Benchmarking and model evaluation:** chosen model is evaluated against relevant benchmarks such as Brier Score and Accuracy Score.

3. **Betting strategy:** models discussed in class are correctly implemented and presented.

4. **Clarity and documentation:** the report is well written, relevant papers are cited, and the layout is well done.

5. **Reproducibility:** your code is reproducible and easy to follow.