Supplement to “The Blown Lead Paradox: Conditional Distribution of the Running Maximum of Binary Forecast Martingales”

Jonathan Pipping-Gamón & Abraham J. Wyner

Department of Statistics, University of Pennsylvania

January 24, 2026

S1 Numerical Benchmarks

This supplement provides numerical benchmarks, analytical examples, and empirical validation of the theoretical results in the main paper, including applications to NFL and NBA win-probability data.

S1.1 Eventual Loser’s Peak Win Probability (Two Teams)

Symmetric Case. For a symmetric matchup with $p_0 = 1/2$, the distribution simplifies to

$$F_{M_\lambda}(x) = \mathbb{P}(M_\lambda \le x) = \begin{cases} 0 & \text{for } 0 \le x < \frac{1}{2}, \\ 2 - \frac{1}{x} & \text{for } \frac{1}{2} \le x < 1, \\ 1 & \text{for } x = 1. \end{cases}$$

So, for example,

$$\begin{aligned} \mathbb{P}(M_\lambda \ge 2/3) &= \frac{1}{2} = 50\%, \\ \mathbb{P}(M_\lambda \ge 0.75) &= \frac{1}{3} \approx 33.3\%, \\ \mathbb{P}(M_\lambda \ge 0.9) &= \frac{1}{9} \approx 11.1\%. \end{aligned}$$

So in a game between two evenly-matched teams, the eventual loser reaches a 66.7% win probability about half the time, 75% about a third of the time, and 90% about one time in nine.

Asymmetric Case. For an asymmetric matchup with $p_0 = 0.75$, the piecewise CDF is

$$F_{M_\lambda}(x) = \mathbb{P}(M_\lambda \le x) = \begin{cases} 0 & \text{for } 0 \le x < 0.25, \\ 1 - \frac{0.25}{x} & \text{for } 0.25 \le x < 0.75, \\ 2 - \frac{1}{x} & \text{for } 0.75 \le x < 1, \\ 1 & \text{for } x = 1. \end{cases}$$

So, for example,

$$\mathbb{P}(M_\lambda \geq 0.5) = \frac{1}{2} = 50\%,$$
$$\mathbb{P}(M_\lambda \geq 0.75) = \frac{1}{3} \approx 33.3\%,$$
$$\mathbb{P}(M_\lambda \geq 0.9) = \frac{1}{9} \approx 11.1\%.$$

So in a 75–25 matchup, the eventual loser still reaches a 50% win probability about half the time, 75% about a third of the time, and 90% about one time in nine.

S1.2 Eventual Winner's Minimum Win Probability (n Teams)

Symmetric Case. In the symmetric n-player case with $p_0^{(i)} = 1/n$, the distribution is

$$F_{M_\omega}(x) = \mathbb{P}(M_\omega \leq x) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{(n-1)x}{1-x} & \text{for } x \in [0, 1/n), \\ 1 & \text{for } x \geq 1/n. \end{cases}$$

For a symmetric three-player game ($n = 3$), we have

$$\mathbb{P}(M_\omega \leq 0.2) = \frac{1}{2} = 50\%,$$
$$\mathbb{P}(M_\omega \leq 0.1) = \frac{2}{9} \approx 22.2\%,$$
$$\mathbb{P}(M_\omega \leq 0.05) = \frac{2}{19} \approx 10.5\%.$$

So in a symmetric three-player game, the winner's minimum win probability falls below 20% about half the time, below 10% two times in nine, and below 5% two times in nineteen.

Asymmetric Case. Consider a three-player game with prior win probabilities $(p_0^{(1)}, p_0^{(2)}, p_0^{(3)}) = (1/6, 1/3, 1/2)$. The distribution of the winner's minimum win probability is

$$F_{M_\omega}(x) = \mathbb{P}(M_\omega \leq x) = \begin{cases} 2 \cdot \frac{x}{1-x} & \text{for } 0 \leq x < 1/6, \\ \frac{1}{6} + \frac{7}{6} \cdot \frac{x}{1-x} & \text{for } 1/6 \leq x < 1/3, \\ \frac{1}{2} + \frac{1}{2} \cdot \frac{x}{1-x} & \text{for } 1/3 \leq x < 1/2, \\ 1 & \text{for } x \geq 1/2. \end{cases}$$

So, for example,

$$\mathbb{P}(M_\omega \leq 0.1) = \frac{2}{9} \approx 22.2\%,$$
$$\mathbb{P}(M_\omega \leq 0.25) = \frac{5}{9} \approx 55.6\%,$$
$$\mathbb{P}(M_\omega \leq 0.4) = \frac{5}{6} \approx 83.3\%.$$

So with prior win probabilities 16.7%, 33.3%, and 50%, the winner's minimum win probability falls below 10% two times in nine, below 25% five times in nine, and below 40% five times in six.

S1.3 Trading Diagnostic: Maximum Win Probability on Losing Trades

Symmetric Case. For a one-sided diagnostic focused on losing trades, take a trade-inception win probability of $p_0 = 0.5$. Conditional on loss, the maximum model-implied win probability has the same CDF as the symmetric two-player case:

$$F_{M|Y=0}(x) = \mathbb{P}(M \le x \mid Y = 0) = \begin{cases} 0 & \text{for } x < 0.5, \\ 2 - \frac{1}{x} & \text{for } 0.5 \le x < 1, \\ 1 & \text{for } x = 1. \end{cases}$$

So, for example,

$$\mathbb{P}(M \ge 2/3 \mid Y = 0) = \frac{1}{2} = 50\%,$$
$$\mathbb{P}(M \ge 0.75 \mid Y = 0) = \frac{1}{3} \approx 33.3\%,$$
$$\mathbb{P}(M \ge 0.9 \mid Y = 0) = \frac{1}{9} \approx 11.1\%.$$

So among losing trades that start at 50%, the model still reaches 66.7% half the time, 75% one third of the time, and 90% one time in nine.

Asymmetric Case. For a one-sided diagnostic focused on losing trades, take a trade-inception win probability of $p_0 = 0.6$. Then the conditional CDF is

$$F_{M|Y=0}(x) = \mathbb{P}(M \le x \mid Y = 0) = \begin{cases} 0 & \text{for } x < 0.6, \\ 1 - \frac{3}{2}\left(\frac{1-x}{x}\right) & \text{for } 0.6 \le x < 1, \\ 1 & \text{for } x = 1. \end{cases}$$

So, for example,

$$\mathbb{P}(M \ge 0.75 \mid Y = 0) = \frac{1}{2} = 50\%,$$
$$\mathbb{P}(M \ge 9/11 \mid Y = 0) = \frac{1}{3} \approx 33.3\%,$$
$$\mathbb{P}(M \ge 0.9 \mid Y = 0) = \frac{1}{6} \approx 16.7\%.$$

So among losing trades that start at 60%, the model still reaches 75% half the time, 81.8% one third of the time, and 90% one time in six.

S2 Empirical Validation Procedures

The theoretical distributions in the main text provide null targets under correct martingale specification. Given n independent realizations in a homogeneous bin (e.g., a narrow range of p_0), compare the empirical CDF $\hat{F}_n$ with the theoretical CDF F_0.

For formal testing, use the *Kolmogorov–Smirnov (K–S) test* (Massey, 1951). The statistic is

$$D_n = \sup_x \left|\hat{F}_n(x) - F_0(x)\right|,$$

and $\sqrt{n}D_n$ converges in distribution to the Kolmogorov distribution under $H_0 : \hat{F}_n = F_0$. Report D_n and its p-value. When comparing multiple p_0 bins, control multiplicity with Bonferroni correction (Dunn, 1961) ($p_i < \alpha/m$) or a false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995). As a secondary descriptive diagnostic, we report the *Kullback–Leibler (KL) divergence* (Kullback and Leibler, 1951) when density estimation is stable.

For visual diagnostics, we recommend:

- *Overlaid CDF plots:* Plot $\hat{F}_n(x)$ and $F_0(x)$ to assess global agreement.
- *Q–Q plots:* Compare empirical and theoretical quantiles; systematic deviations reveal location, scale, or tail mismatches.

Practical considerations:

- *Binning:* Use narrow p_0 bands (e.g., $[0.49, 0.51]$) to maintain approximate homogeneity while preserving sample size.
- *Finite-sample effects:* For small N, the discrete-time bounds may be more appropriate; the Brownian formulas improve as N grows.
- *Model misspecification:* Systematic deviations from F_0 indicate over/underconfidence or violations of the martingale assumptions, motivating model refinement.

S3 Empirical Validation on NFL and NBA Data

S3.1 Data Sources and Preprocessing

We validate the theoretical distributions using ESPN win-probability and play-by-play data for NFL and NBA regular-season games from 2018–2024 (ESPN, 2026). NFL data are pulled via `espnscrapeR` (Mock, 2025), and NBA data via `hoopR` (Gilani, 2023). For each game, we define the favored team's starting probability p_0 as the first reported win probability in the ESPN feed, and compute M_λ as the maximum win probability attained by the eventual loser over the game. We exclude ties and drop games with missing scores or missing win-probability series.

S3.2 Empirical CDFs

Games are grouped into centered bins of width 0.05 with centers at $p_0 \in \{0.50, 0.55, \ldots, 0.95\}$. We pool all seasons within each league, estimate the empirical CDF of M_λ within each bin, and compare it against the theoretical CDF for the matching p_0. Bins with fewer than 100 games are omitted. Figures 1 and 2 show CDF overlays for $p_0 \in \{0.50, 0.60, 0.70, 0.80\}$.

S3.3 Diagnostic Tables

For each bin, Tables 1 and 2 report, in order, the starting win probability p_0, the number of games n, the KL divergence between the empirical and theoretical distributions, the K–S statistic D_n, its p-value, and a Bonferroni rejection indicator (NFL: $\alpha = 0.05/7$; NBA: $\alpha = 0.05/9$).

In these data, the NFL bins exhibit few Bonferroni rejections (only $p_0 = 0.80$), consistent with a reasonably calibrated model at the bin level, whereas the NBA bins reject across the board, indicating systematic deviations from the theoretical distribution and significant model miscalibration.

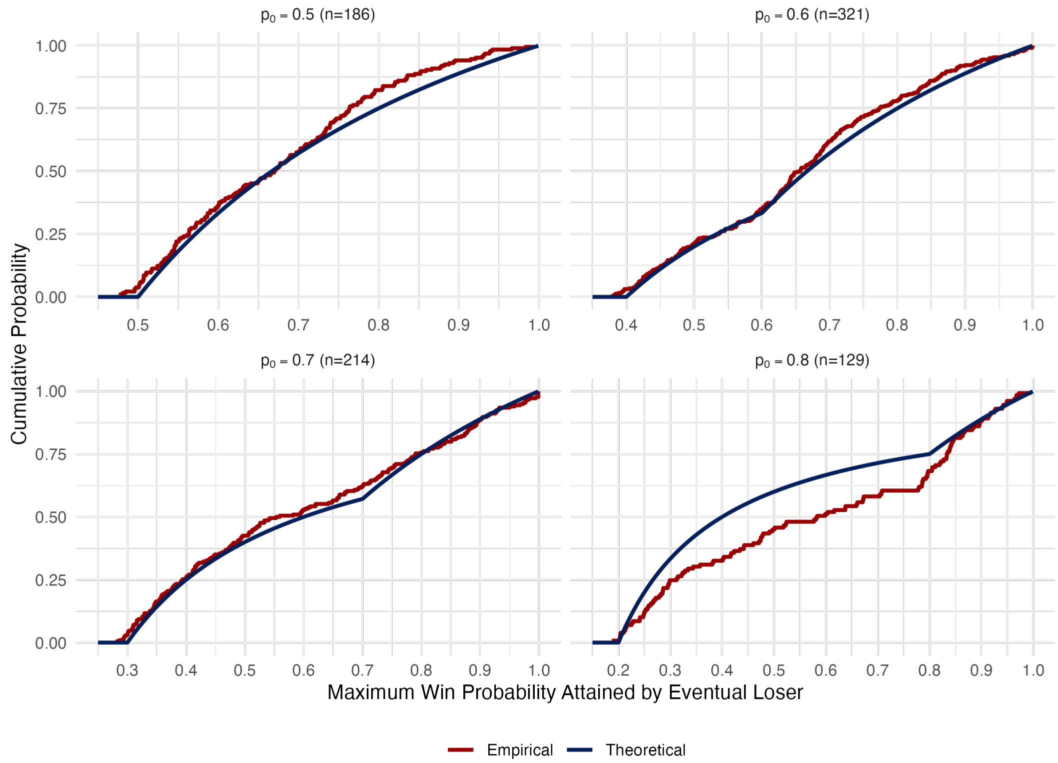

Figure 1: NFL CDF overlays by p_0 bin.

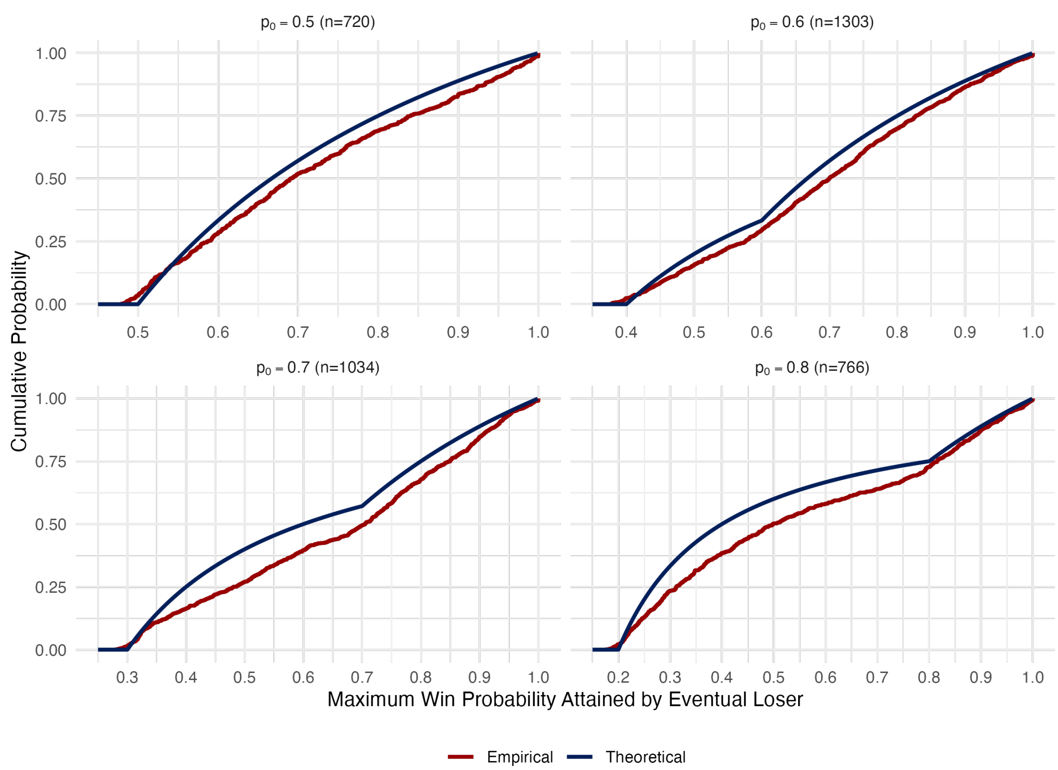

Figure 2: NBA CDF overlays by p_0 bin.

Table 1: NFL diagnostic summary by p_0 bin.

p_0	n	KL	K–S D_n	p-value	Reject
0.50	186	0.916	0.0806	0.1779	×
0.55	352	0.685	0.0827	0.0162	×
0.60	321	0.740	0.0602	0.1951	×
0.65	304	0.112	0.0768	0.0556	×
0.70	214	0.793	0.0553	0.5286	×
0.75	179	1.111	0.0676	0.3873	×
0.80	129	0.314	0.1800	0.0005	✓

Table 2: NBA diagnostic summary by p_0 bin.

p_0	n	KL	K–S D_n	p-value	Reject
0.50	720	0.872	0.0712	0.0014	✓
0.55	1393	0.499	0.0854	0.0000	✓
0.60	1303	0.588	0.0755	0.0000	✓
0.65	1188	0.419	0.0962	0.0000	✓
0.70	1034	0.404	0.1332	0.0000	✓
0.75	953	0.627	0.1200	0.0000	✓
0.80	766	0.540	0.1251	0.0000	✓
0.85	520	0.719	0.1658	0.0000	✓
0.90	286	0.793	0.2166	0.0000	✓

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, 56(293):52–64.

ESPN (2026). ESPN Play-by-Play and Win Probability Data. Accessed 2026-01-22.

Gilani, S. (2023). *hoopR: Access Men's Basketball Play by Play Data*. R package version 2.1.0.

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78.

Mock, T. (2025). *espnscrapeR: Scrapes Or Collects NFL Data From ESPN*. R package version 0.8.0.