

Window-Adjusted WAR (wWAR)

A Team-Specific Framework for Prospective MLB Player Value Under Competitive Windows

Michael Pilger

Working mini-paper | September 8, 2026

Research question: How much should the same projected WAR be worth to different MLB franchises when championship windows, long-run franchise strength, and postseason leverage differ?

ABSTRACT

Wins Above Replacement (WAR) is intentionally context-neutral: a projected win above replacement has the same numerical value regardless of the team receiving it. This paper develops Window-Adjusted WAR (wWAR), a prospective, team-specific extension that asks how efficiently a player's projected WAR converts into championship probability for a particular MLB franchise over a competitive window. The framework combines temporally validated player WAR forecasts, a learned transition from current-team state toward persistent franchise structure, correlated five-year Monte Carlo simulation, and a fixed league reference for the marginal championship value of one WAR. It also places the baseball-value metric inside an acquisition framework in which contract cost, Competitive Balance Tax accounting, contract length, and trade acquisition cost are treated downstream from wWAR rather than embedded inside player performance. Historical validation shows that the player model beats WAR persistence at every forecast horizon, while endpoint ablation indicates that championship forecasts should become structural earlier than playoff forecasts. In the final 2026-2030 Shohei Ohtani case study, the same 19.81 projected WAR produces 29.32 wWAR for the Yankees, 27.74 for the Dodgers, and 11.06 for the Rockies. A separate 100,000-universe independent-seed replication confirms the Yankees-Dodgers ordering and overwhelmingly reproduces the elite-franchise versus Colorado gap. Finally, the paper specifies a reproducible any-player assignment procedure: given an annual projected-WAR path and a current team, the frozen engine can calculate wWAR for the player's actual club or every potential destination. The resulting framework connects player projection, competitive-window timing, contract economics, transaction valuation, and league-wide competitive balance without redefining descriptive WAR itself.

Keywords: WAR, competitive windows, MLB, player valuation, championship probability, contracts, trades, Competitive Balance Tax, Monte Carlo simulation, franchise strength, reproducibility

1. Introduction

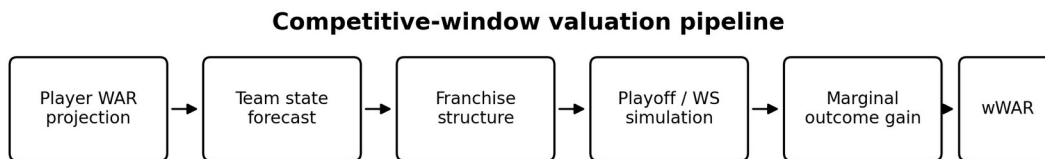
WAR is one of baseball analytics' most useful common currencies because it compresses batting, pitching, baserunning, fielding, positional adjustment, and replacement level into wins. That construction is intentionally portable across players and teams. FanGraphs describes WAR as an estimate of the wins a player contributes above a freely available replacement alternative, not as a complete model of every downstream consequence of those wins (Slowinski, 2012). The portability is a strength for player comparison, but it also creates a gap for prospective decision-making.

A win does not have the same strategic consequence in every organizational context. An elite player added to a team near or inside a championship window can convert additional regular-season strength into playoff byes, deeper October runs, and World Series probability. The identical player placed on a weaker franchise may create substantial baseball production while barely moving championship odds. Over longer horizons, the relevant context is not only the present roster. Some franchises repeatedly build competitive teams, while others cycle through shorter and less frequent windows. A useful prospective value statistic therefore needs both a current-state component and a franchise-specific long-run component.

The motivating example is Shohei Ohtani. Standard WAR can estimate Ohtani's on-field production, but it cannot answer whether that production is more valuable to the Dodgers, Yankees, Rockies, or another destination. This paper builds a model in which Ohtani's projected WAR is held fixed and only the destination environment changes. The resulting statistic, Window-Adjusted WAR (wWAR), scales raw projected WAR by the player's simulated marginal effect on expected championships relative to a fixed league reference.

The objective is not to replace descriptive WAR. wWAR is a forward-looking decision statistic. It is designed to answer a narrower question: given a player's projected MLB production and a specified control horizon, how much championship-adjusted baseball value does that player create for each possible team? A secondary metric, playoff-adjusted WAR (pWAR), uses the same architecture but values marginal playoff appearances rather than championships.

This distinction has direct business consequences. A front office does not acquire WAR in isolation; it acquires a player at a salary, for a specific number of control years, under a payroll and Competitive Balance Tax environment, and sometimes at the additional cost of prospects, draft capital, international bonus-pool space, or other transaction assets. Two clubs can therefore rationally place different bid values on the same player even when they agree perfectly on the player's raw WAR forecast. wWAR supplies the baseball-outcome side of that bid. Contract and acquisition costs can then be layered on afterward to estimate whether the transaction creates positive team-specific surplus.



Hybrid short-run signal fades toward structural franchise strength; the same player is inserted into each destination under common random numbers.

Primary endpoint: expected World Series titles. Secondary endpoint: expected playoff appearances.

Figure 1. Conceptual architecture of the wWAR model. Player production is forecast independently, then translated through team state, franchise structure, and a joint postseason simulator into marginal championship value.

1.1 Research hypotheses

The model is built around four testable claims rather than an arbitrary multiplier.

- H1. Current team strength should dominate near-term forecasts, but persistent franchise structure should receive increasing weight as the forecast horizon lengthens.
- H2. The same projected WAR should generate different marginal playoff and championship effects across destination teams.
- H3. Championship leverage and playoff leverage should not be identical objective functions. Teams can be near the playoff margin without being in the highest championship-leverage state.
- H4. A fixed league reference should be stable enough across seasons to make wWAR comparable across teams and, eventually, across players.

2. Data and Study Design

The project combines player-season production, team-season performance, postseason series outcomes, roster identity, and payroll or market information. Player WAR histories are built from FanGraphs-style season records covering the modern modeling window, with explicit treatment of the shortened 2020 season. Historical team and postseason outcomes are drawn from the SABR Lahman Baseball Database through 2025. The Lahman database provides team records, franchise identifiers, and postseason series results, and is maintained by SABR from the historical database created by Sean Lahman. Payroll data used for the structural-resource tests come from the Wharton Sports Analytics and Business Initiative teaching dataset, which covers MLB payrolls from 1998 through 2023.

The main modeling period is deliberately temporal. Player forecasting begins with training data in 2012 and evaluates future seasons without random train-test mixing. Team-strength and probability models are also tested using sequential out-of-sample target seasons. This design matters because a random split would leak future league environments, player aging patterns, and franchise trajectories into earlier predictions.

The current postseason simulator implements MLB's 12-team format: six qualifiers per league, three division champions, three wild cards, first-round byes for the top two seeds, best-of-three Wild Card Series, best-of-five Division Series, and best-of-seven League Championship and World Series rounds. Historical postseason calibration uses multi-game Division Series, League Championship Series, and World Series from 1995 through 2025; Wild Card rounds are excluded from the strength calibration because their format changed repeatedly.

3. Methods

3.1 Player WAR forecasting

The player layer forecasts expected MLB WAR one through five seasons ahead. Candidate systems included zero-WAR and persistence baselines, weighted recent WAR, direct Ridge and gradient-boosting regressions, and hurdle models. The hurdle structure estimates both the probability that a player appears in MLB in the target season and expected WAR conditional on appearing, then multiplies the two quantities. This is particularly important at long horizons because nonappearance is a real baseball outcome rather than missing data.

Model selection uses pooled temporal out-of-sample mean absolute error (MAE), with root mean squared error (RMSE) breaking practical ties. Hitters select a HistGradientBoosting hurdle model at one year and Ridge hurdle models at longer horizons; pitchers select Ridge hurdle models at all five horizons. Statcast features were tested but were not retained because their matched-window gains did not reliably exceed the best longer-history core models.

Season	Projected WAR
2026	5.70
2027	4.39
2028	3.96
2029	3.16
2030	2.59

Table 1. Ohtani production path used in the final case study. Five-year expected WAR = 19.81.

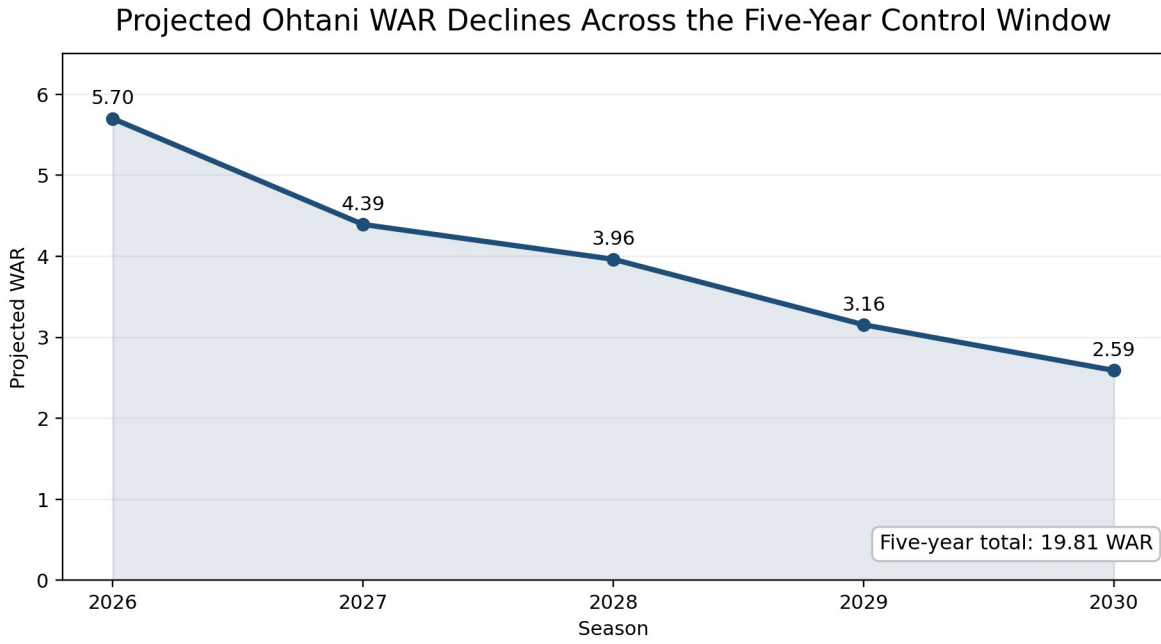


Figure 3. Ohtani's projected WAR path used in the case study. The model values the timing of production, not only the five-year total, because early WAR interacts with a different competitive environment than later WAR.

3.2 Team strength: current state versus franchise structure

Future team strength is modeled as a horizon-specific blend of two independently validated signals. CurrentState summarizes the present competitive team and is strongest near term. StructuralStrength represents persistent franchise quality estimated from historical out-of-sample franchise performance. The final team-strength forecast is:

$$\text{FinalStrength}_{j,h} = \lambda_h \text{CurrentState}_{j,h} + (1 - \lambda_h) \text{StructuralStrength}_{j,h} \quad (1)$$

The blend weights are learned only from historical out-of-sample predictions and then constrained to decay with horizon. The estimated current-state weights are 0.773, 0.486, 0.206, 0.013, and 0.013 from horizons one through five. In other words, the model does not impose a universal five-year championship window. It learns that current roster information decays rapidly and long-run franchise structure dominates later seasons.

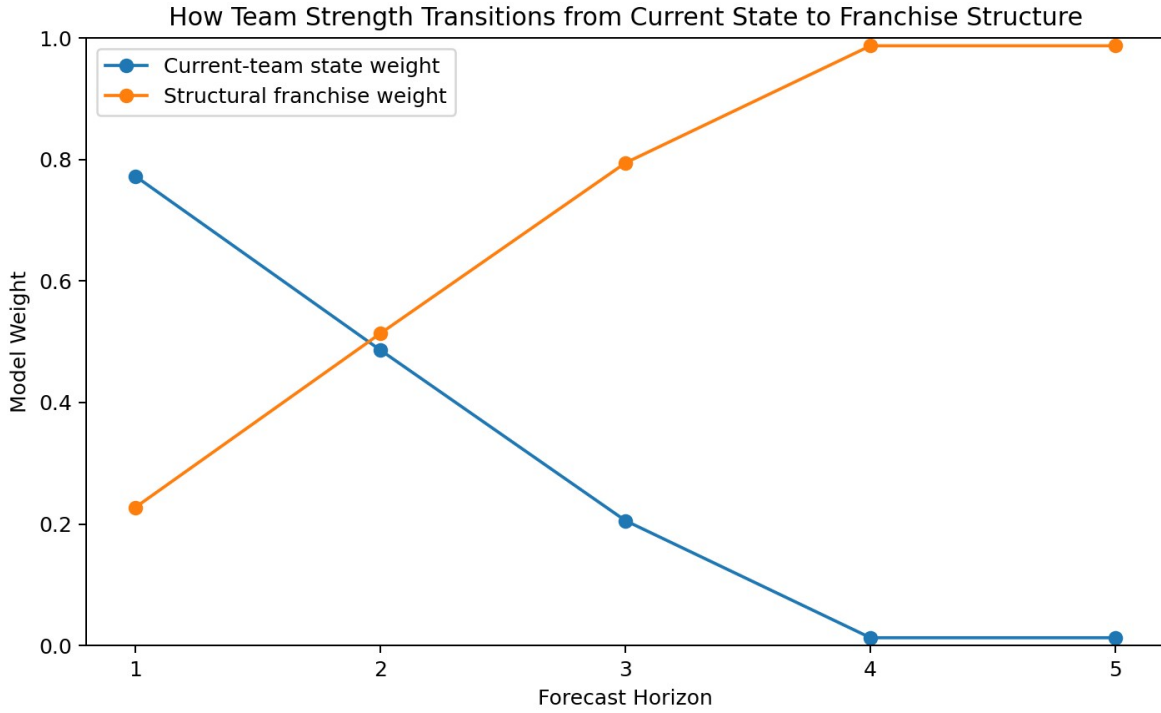


Figure 2. Learned transition from current-team state to structural franchise strength. By years four and five, the forecast is almost entirely structural.

Payroll was tested as a structural resource variable. Payroll capacity is highly persistent, but it did not improve future-win MAE after historical franchise winning was included. It is therefore retained for downstream economic analysis rather than forced into the team-strength forecast. This is important conceptually: richer organizations can receive higher long-run equilibrium estimates if that advantage manifests in persistent winning, but market size is not hard-coded as a direct win bonus.

3.3 Joint five-year playoff and championship simulation

The team-strength paths are converted into postseason probabilities with a joint Monte Carlo model. Regular-season uncertainty is decomposed into latent team-quality uncertainty and within-season record noise. Cross-season latent franchise shocks follow an AR(1) process, with persistence estimated from historical structural forecast residuals. The implied latent persistence is approximately 0.68, allowing unexpectedly strong or weak organizational states to carry into subsequent seasons rather than resetting independently each year.

Postseason series are modeled with a Bradley-Terry paired-comparison framework (Bradley and Terry, 1952). The mapping from latent team strength to series win probability is calibrated on historical Division Series, League Championship Series, and World Series outcomes. Parameter uncertainty is propagated by drawing the postseason strength coefficient from a postseason-year cluster bootstrap for each simulated five-year universe.

The final production simulation uses common random numbers. For a baseline universe and the corresponding player-added counterfactual, latent franchise shocks, regular-season random draws, postseason coefficient draws, and postseason random uniforms are identical. Only the player insertion changes. This paired design sharply reduces Monte Carlo noise in the marginal treatment effect.

Horizon	Championship endpoint	Playoff endpoint
H1	Hybrid	Hybrid
H2	Structural	Hybrid
H3	Structural	Structural
H4	Structural	Structural
H5	Structural	Structural

Table 2. Final endpoint-specific team-strength schedules, frozen before the production run.

3.4 Player insertion and double-counting control

For each candidate destination, the player's projected WAR path is added to that team. The same total number of wins is removed evenly across the other 29 clubs so that the league remains approximately zero-sum. Because the short-run Hybrid forecast already partially reflects players on their 2025 clubs, the model first removes the embedded share of the player from the current team for Hybrid horizons only. Structural baselines are treated as franchise-level and are not player-specific.

$$\text{EmbeddedWAR}_{p,h} = \lambda_h \text{WAR}_{p,h} \quad (\text{Hybrid horizons only}) \quad (2)$$

3.5 Fixed-reference wWAR

A central problem in early versions of the statistic was normalization. If every player is divided by that same player's average championship gain across teams, the mean wWAR is forced to equal raw WAR and cross-player comparisons become distorted when championship conversion is nonlinear. The final specification instead uses a player-independent fixed reference.

For each horizon h , the simulator adds the same standardized +5 WAR dose to every franchise and estimates the league-average marginal championship gain per WAR, κ_h^{WS} . For player p , the neutral expected championship value of the player's WAR path is:

$$R_p^{WS} = \sum_{h=1}^H \text{WAR}_{p,h} \kappa_h^{WS} \quad (3)$$

For destination j , the championship multiplier and wWAR are then:

$$M_{p,j}^{WS} = \frac{\Delta \mathbb{E}[\text{Titles} \mid p \rightarrow j]}{R_p^{WS}} \quad (4)$$

$$\text{wWAR}_{p,j} = \text{WAR}_{p,\text{total}} M_{p,j}^{WS}$$

(5)

$$\text{pWAR}_{p,j} = \text{WAR}_{p,\text{total}} \frac{\Delta \mathbb{E}[\text{PlayoffApps} \mid p \rightarrow j]}{\sum_h \text{WAR}_{p,h} \kappa_h^{PO}}$$

The secondary pWAR statistic is defined identically using horizon-specific league-average marginal playoff-appearance gains. The fixed reference makes a neutral multiplier equal to 1.0 without forcing every individual player's mean destination value to equal raw WAR.

3.6 Generalized wWAR assignment for any MLB player

The Ohtani example is a case study, not a player-specific definition. The final statistic is portable to any MLB player once the analyst supplies an annual projected-WAR path over the desired control horizon and identifies the player's current club for short-run double-counting control. Let $j(p)$ denote the club on which player p is being evaluated. The player's assigned value is simply the destination-specific output for that club:

$$\text{Assigned wWAR}_p = \text{wWAR}_{p, j(p)}$$

A reproducible scoring run therefore requires only four player-specific inputs: identity, current team, annual projected WAR, and the evaluation horizon. All team-strength paths, structural baselines, postseason calibration, endpoint schedules, and fixed-reference conversion rates are frozen model inputs. A user can request the player's wWAR on the actual team or evaluate all 30 destinations without refitting the model. This distinction is essential for reproducibility: assigning a new player is a scoring operation, not a new research phase.

For cross-player comparison, the annual timing of WAR must be preserved. Two players with the same five-year total can receive different wWAR because a front-loaded WAR path interacts more heavily with the current competitive window, while a back-loaded path interacts more heavily with structural franchise strength. The companion reproducibility package delivered with this paper implements this input schema directly.

4. Historical Validation and Robustness

4.1 Player forecast validity

The Phase 3 player forecast clearly beats current-WAR persistence at every horizon. Across the complete historical player population, MAE improvements versus persistence rise from 0.204 WAR at one year to 0.384 WAR at five years. The zero-WAR baseline becomes difficult to beat for the full fringe-player population at long horizons because many marginal players leave MLB. This does not invalidate the intended wWAR population: for players projected at 2+ WAR, the model beats zero WAR at every horizon, with mean MAE improvements of 1.24, 1.05, 0.78, 1.27, and 0.75 WAR from horizons one through five.

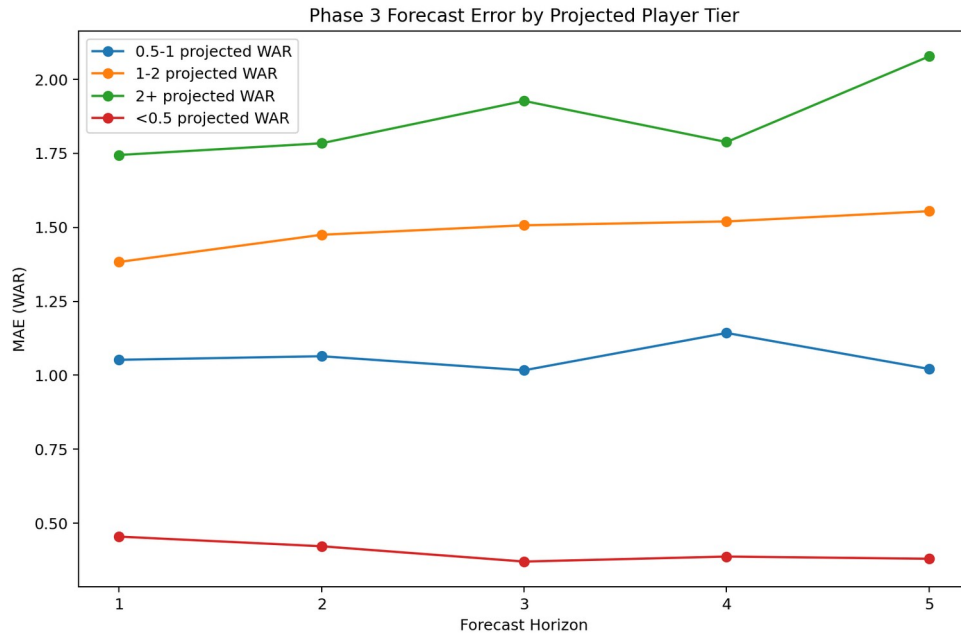


Figure 4. Player forecast MAE by projected-value tier. Higher-value players have greater absolute outcome variance, but the model retains meaningful predictive advantage over naive baselines for the 2+ WAR population.

4.2 Team-strength and probability backtests

The hybrid team model improves one-year historical win forecasts relative to CurrentState, Structural, and a league-mean baseline. One-year MAE is 9.33 wins for Hybrid versus 9.52 for CurrentState, 10.15 for Structural, and 10.71 for the 81-win baseline. At two years, Hybrid remains best at 9.84 wins. By year three, Structural slightly outperforms Hybrid, consistent with the learned decay of current-team information.

The probability layer also clears a meaningful benchmark. At one year, the Hybrid playoff model posts a Brier score of 0.2065 versus 0.2385 for the naive format-based probability, corresponding to 13.4% Brier skill; AUC is 0.725. Skill declines with horizon, reaching roughly 5.4% at two years and near zero by three years. This degradation is a feature rather than a defect: the model explicitly becomes less confident in team-specific near-term differences as the horizon expands.

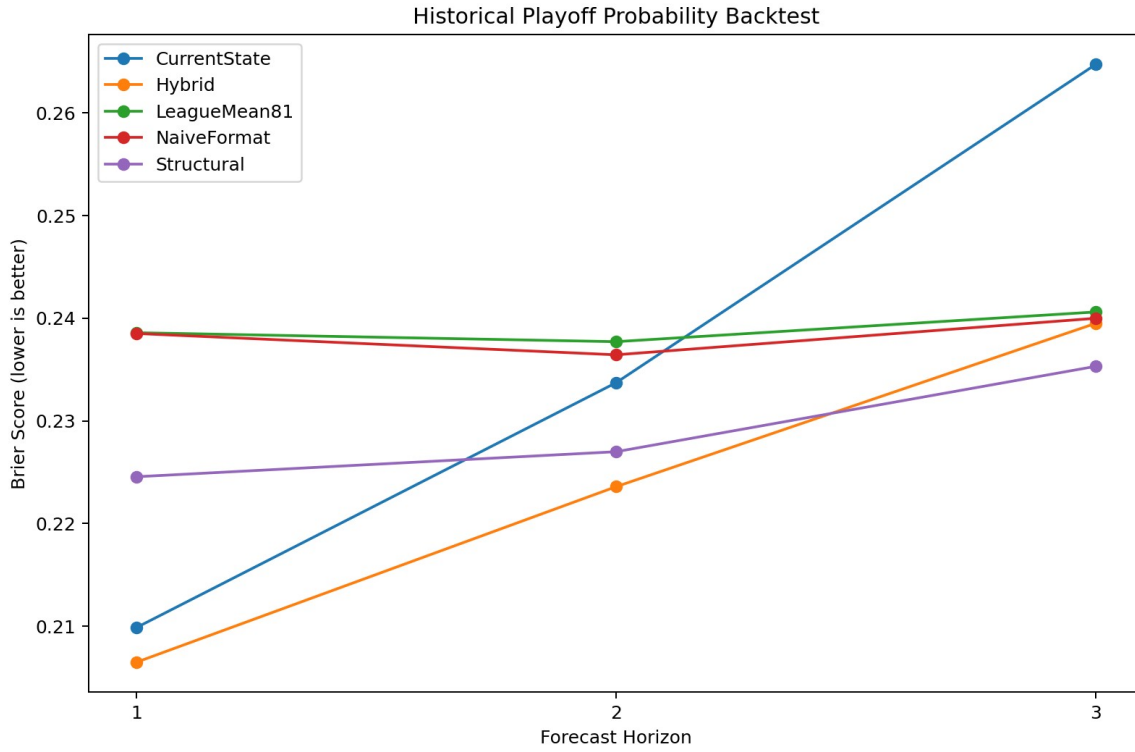


Figure 5. Historical playoff-probability backtest. Lower Brier score is better; the Hybrid model is strongest at the one-year horizon and loses advantage as the horizon grows.

4.3 Endpoint ablation and simplification tests

Historical endpoint ablation showed that the best information schedule differs by objective. For playoff qualification, Hybrid remains preferred at horizons one and two, then Structural is better at horizon three. For World Series probability, Hybrid is slightly better at horizon one, but Structural has a lower H2 Brier score (0.03173 versus 0.03247) and the leave-one-season-out blend assigns 0% weight to Hybrid at H2. The final championship schedule therefore becomes structural beginning in year two.

The full simulator was also challenged by simple probability heuristics. A two-probability derivative approximation using baseline World Series and playoff probabilities explains 77.8% of historical out-of-sample marginal championship-leverage variance and 84.2% of the forward Ohtani wWAR variation across teams. That is too strong to dismiss, but too weak to replace the simulator: the remaining error is concentrated in several high-leverage destinations where league, division, and bracket interactions matter most.

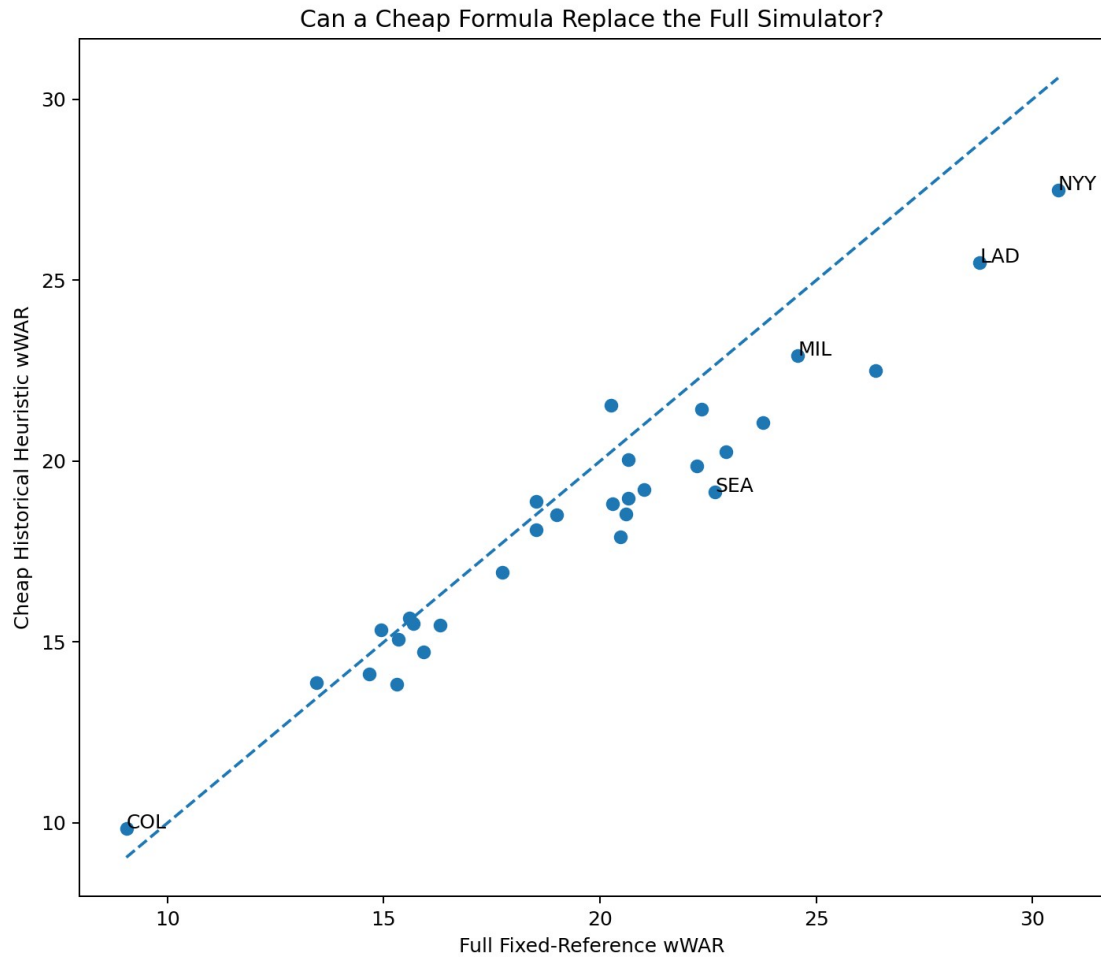


Figure 6. Cheap probability heuristic versus full fixed-reference wWAR in the development Ohtani experiment. The approximation is strong but leaves material destination-specific error.

4.4 Fixed-reference stability

The fixed reference is stable enough to support normalization. Excluding the unusual 2020 season, the historical 2021-2025 mean championship reference is 0.00435 expected titles gained per WAR with a coefficient of variation of 6.95%. The forward 2026-2030 mean is 0.00482 with a coefficient of variation of 5.67%, only 10.7% higher than the post-2020 historical mean. The analogous playoff reference differs by 9.8%.

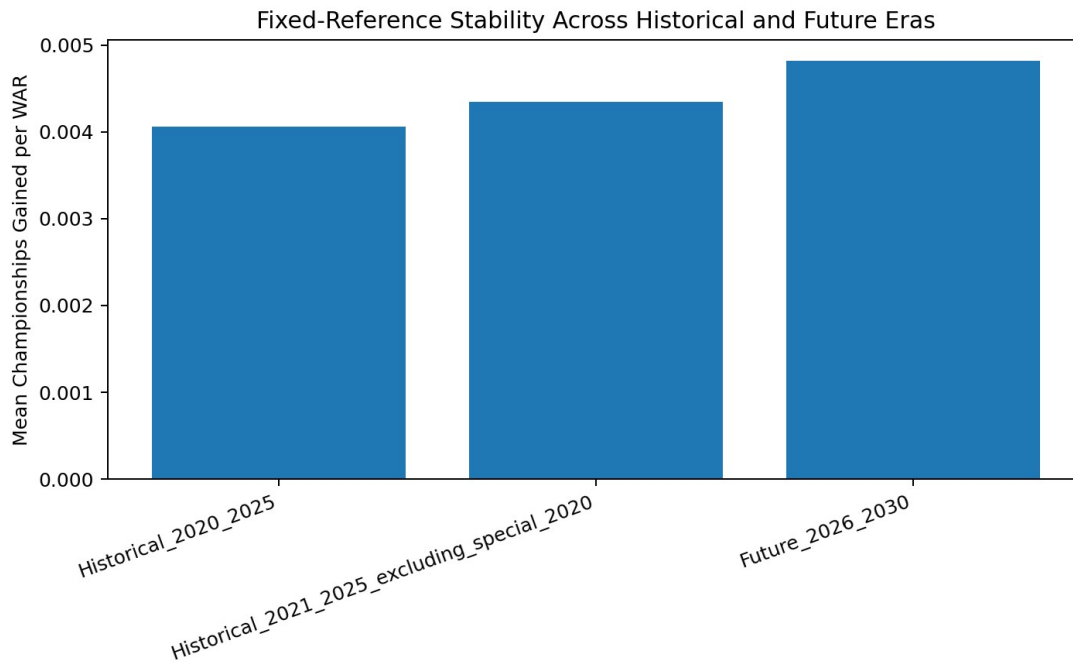


Figure 7. Fixed-reference stability across historical and forward eras. Excluding the special 2020 postseason regime materially reduces the historical-to-forward difference.

5. Final Ohtani Case Study

The final case study holds Ohtani's expected 2026-2030 production fixed at 19.81 WAR for every destination. The only changing variable is team context. The production run uses 50,000 paired player-counterfactual universes with the final endpoint schedules frozen in advance. The dashed vertical line in Figure 7 marks raw WAR; bars to the right indicate championship amplification and bars to the left indicate championship discount relative to the fixed league reference.

The Same 19.81 WAR Has Very Different Championship Value by Destination

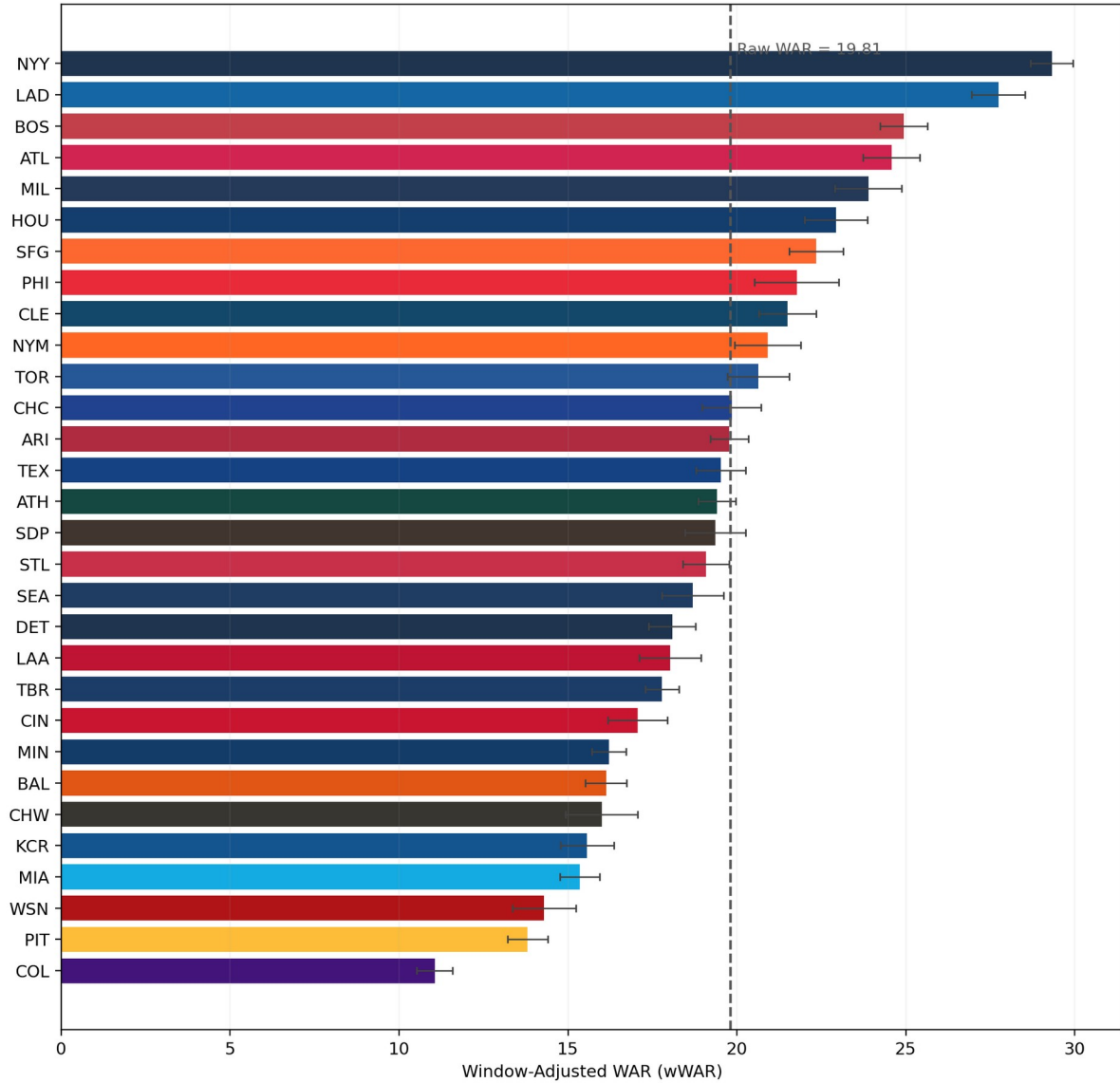


Figure 8. Final team-specific wWAR for Shohei Ohtani, 2026-2030. The same 19.81 projected WAR is amplified or discounted by destination-specific championship leverage. Error bars are Monte Carlo intervals from the frozen 50K production run.

Team	Raw WAR	wWAR	wWAR 95% MC CI	Multiplier	Delta exp. titles	Delta P(>=1 WS)	pWAR
NYN	19.81	29.32	28.70-29.95	1.481x	0.136	9.41 pp	18.95
LAD	19.81	27.74	26.95-28.54	1.401x	0.129	8.91 pp	19.47
COL	19.81	11.06	10.52-11.59	0.558x	0.051	4.32 pp	15.37

Table 3. Final production estimates for the motivating destinations.

The Yankees produce the highest final wWAR at 29.32, followed by the Dodgers at 27.74. Colorado produces 11.06. Relative to the neutral reference, the championship multipliers are 1.481 for New York, 1.401 for Los Angeles, and 0.558 for Colorado. The same 19.81 raw WAR therefore receives a roughly 48% championship-context premium in New York, a 40% premium in Los Angeles, and a 44% discount in Colorado.

The result is also interpretable directly in championship probability. Adding Ohtani increases five-year expected championships by 0.136 for the Yankees, 0.129 for the Dodgers, and 0.051 for the Rockies. The modeled probability of at least one World Series title rises by 9.41 percentage points in New York, 8.91 points in Los Angeles, and 4.32 points in Colorado.

Paired production-run contrasts are large for the elite-versus-Colorado comparisons:

LAD - COL: 16.69 wWAR (MC interval 15.72 to 17.66).

NYN - COL: 18.27 wWAR (MC interval 17.53 to 19.00).

NYN - LAD: 1.58 wWAR (MC interval 0.52 to 2.64).

5.1 Championship value versus playoff value

The secondary pWAR results demonstrate that championship and playoff objectives are not interchangeable. Ohtani's pWAR is 19.47 for the Dodgers and 18.95 for the Yankees even though wWAR is higher for New York. In the paired production run, pWAR_NYY - pWAR_LAD is -0.51 with a Monte Carlo interval from -0.82 to -0.21. The independent-seed replication produces the same direction.

This divergence is substantively plausible. A team already extremely likely to reach October can have diminishing marginal playoff value while still converting additional wins efficiently into deeper postseason success. pWAR is therefore best interpreted as marginal value toward playoff qualification; wWAR is marginal value toward championships.

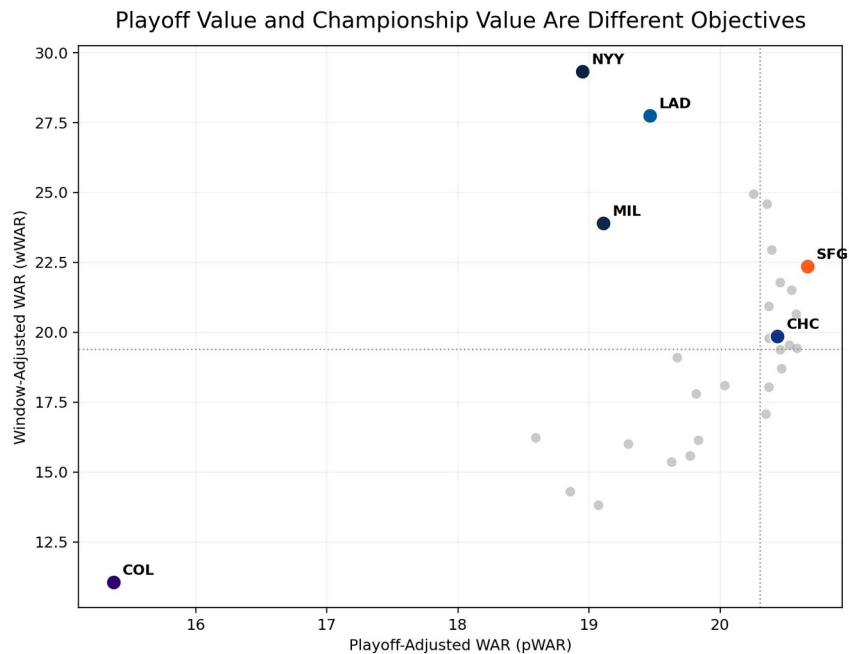


Figure 9. pWAR and wWAR measure distinct forms of marginal value. Teams near the playoff margin can score highly on pWAR without being the highest-championship-leverage destination, while elite championship environments can rank much higher on wWAR.

Championship Leverage Decays as the Forecast Moves Away from the Current Window

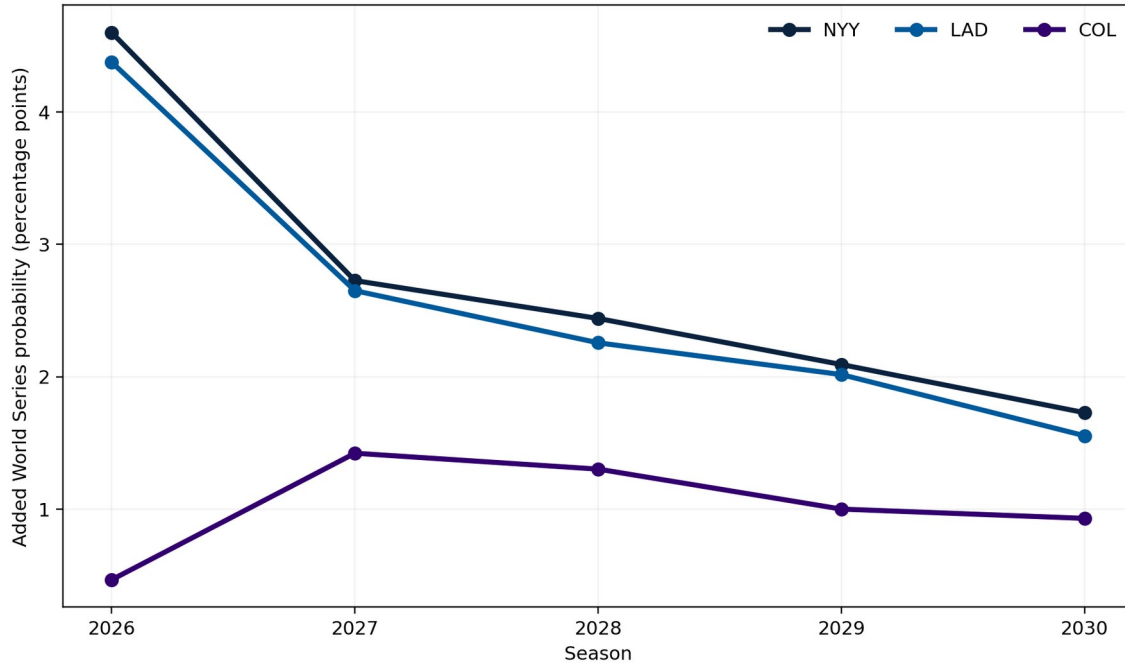


Figure 10. Annual championship impact of Ohtani for the Yankees, Dodgers, and Rockies. Elite-team leverage is highest immediately, while the gap narrows as the model transitions toward long-run structural franchise strength.

5.2 Independent-seed replication

Because the 25,000- and 50,000-universe development runs changed the exact Yankees-Dodgers ordering, the final specification was subjected to a separate, precommitted replication test. The model was frozen, the completed neutral reference denominator was held fixed, and 100,000 completely new universes were generated across 20 independent batches for only the Yankees, Dodgers, and Rockies. The decision rule was established before observing the new result: the Yankees-Dodgers ordering would be considered separable only if the independent 95% interval excluded zero and matched the 50K sign.

The replication passes that rule. Replication wWAR is 28.81 for New York, 26.95 for Los Angeles, and 11.04 for Colorado. The independent NYY - LAD contrast is +1.86 wWAR with a 95% batch-t interval from +0.67 to +3.04. LAD - COL is +15.92 and NYY - COL is +17.78, both overwhelmingly positive. The 100,000-universe estimates differ from the 50K production wWAR values by only -1.75% for NYY, -2.85% for LAD, and -0.19% for COL.

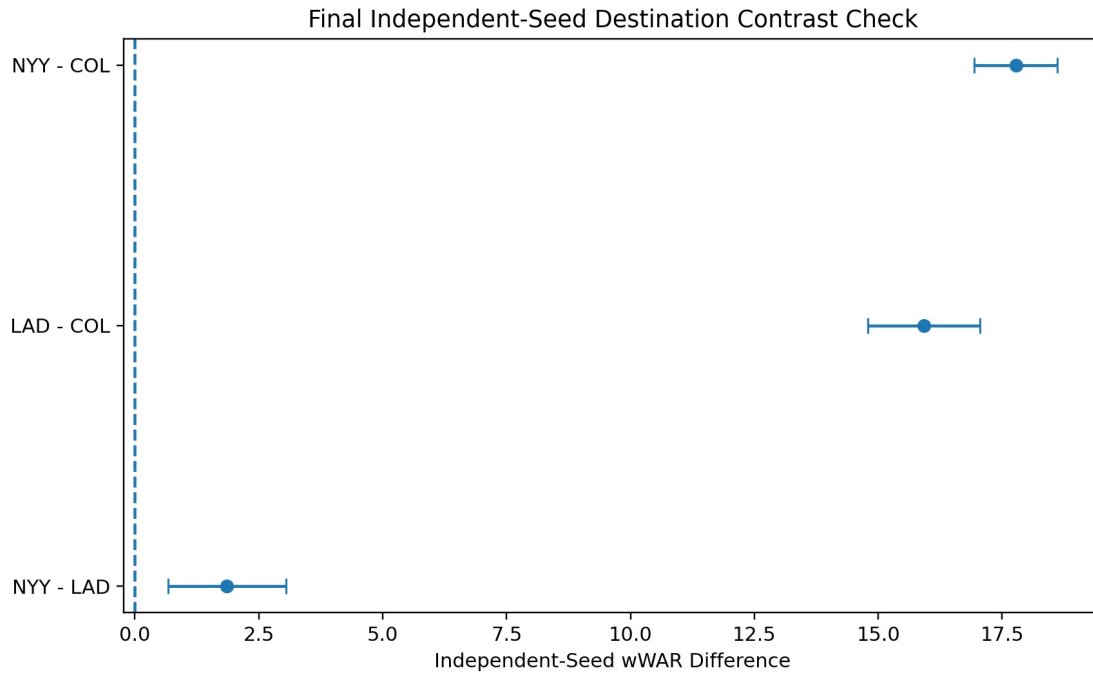


Figure 11. Independent-seed replication of the three key destination contrasts. All intervals exclude zero, including the more delicate Yankees-Dodgers comparison.

6. Contract Economics and Transaction Value

A front office acquires a player as both a baseball asset and a contractual liability. wWAR deliberately separates those concepts. The baseball model first estimates how much championship value the player creates for the destination. Only after that value is estimated are salary, Competitive Balance Tax (CBT) accounting, contract length, deferrals, and acquisition costs applied. This avoids contaminating the player-performance forecast with team-specific accounting choices while still allowing wWAR to inform real transactions.

6.1 Salary, CBT accounting, and deferred compensation

For MLB clubs, cash salary and CBT cost are not the same object. Under the 2022-2026 collective bargaining agreement, CBT payroll is generally based on average annual contract value and benefits rather than the cash paid in a given season. The 2026 base CBT threshold is \$244 million, with escalating rates for repeat payors and additional surcharges for larger overages. This matters because a transaction can be inexpensive in near-term cash while still consuming substantial CBT capacity.

Ohtani is an unusually clear example. His contract pays \$2 million in annual cash during the playing term while most compensation is deferred, but the deal still creates a much larger annual CBT accounting charge. Figure 12 separates projected on-field production value, cash payment, and CBT charge across the paper's five-year control window. The three series answer different business questions: liquidity, tax/payroll capacity, and expected baseball production.

Cash Flow, CBT Accounting, and On-Field Production Are Different Economic Objects

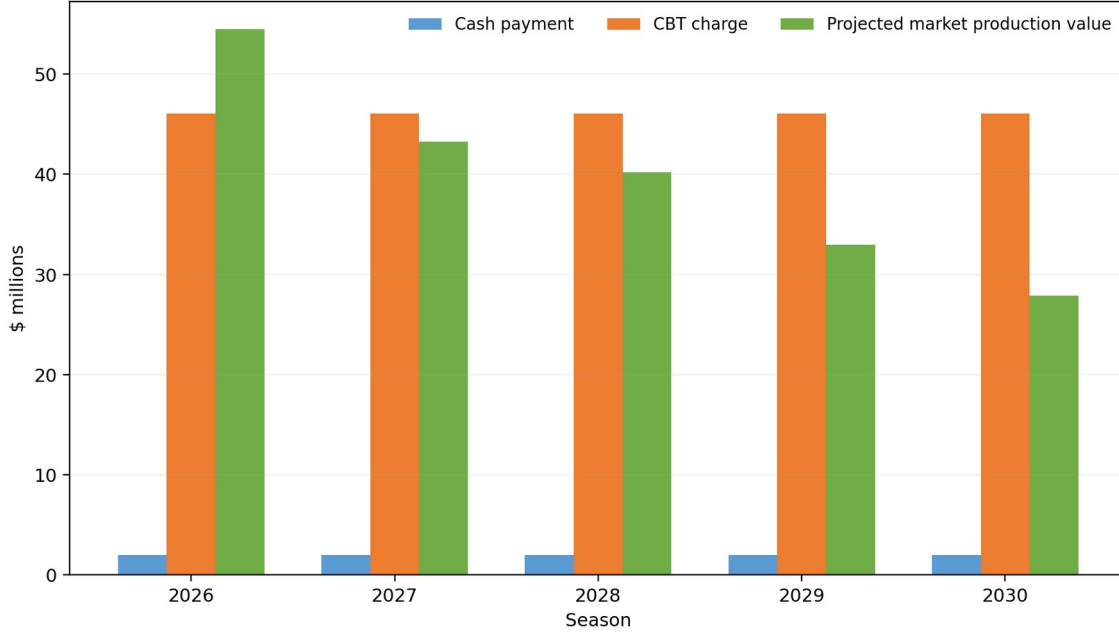


Figure 12. Ohtani cash payment, CBT charge, and projected market production value during 2026-2030. Deferred cash can materially change liquidity without eliminating CBT or economic opportunity cost.

$$\text{ContractCost}_{p,j} = \text{PV}(\text{Salary}) + \text{PV}(\text{CBT}) + \text{OtherAcquisitionCosts}$$

6.2 Team-Adjusted Asset Value (TAAV)

The paper translates wWAR into a downstream financial scenario measure called Team-Adjusted Asset Value (TAAV). Let m denote an assumed market price of one WAR. The same contract can have different TAAV across clubs because wWAR is destination-specific even when salary is fixed.

$$\text{TAAV}_{p,j}(m) = m \cdot \text{wWAR}_{p,j} - \text{ContractCost}_{p,j} \quad (6)$$

The market price of WAR is itself uncertain, so a single TAAV point estimate would create false precision. The final analysis therefore reports a range of market-rate scenarios. At approximately \$8.54 million per WAR, the Yankees generate about +\$20.1 million of TAAV, the Dodgers about +\$6.6 million, and Colorado about -\$135.9 million. At roughly \$10.04 million per WAR, both NYY and LAD are strongly positive while Colorado remains negative. Even at the 97.5th-percentile market-rate scenario of about \$12.29 million per WAR, Colorado remains approximately -\$94.5 million.

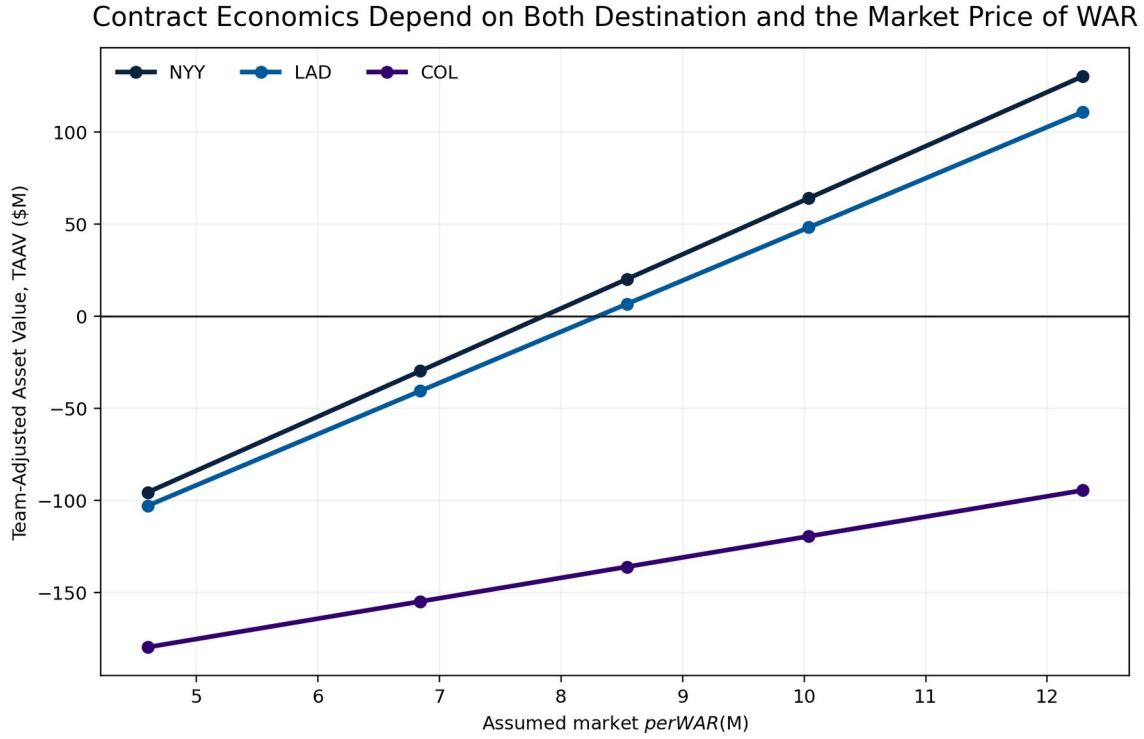


Figure 13. Team-adjusted asset value under alternative market prices of WAR. The break-even market price differs materially by destination because the same contract purchases different championship-adjusted baseball value.

6.3 Free agency, trades, and extensions

The economic logic extends beyond free agency. In a free-agent signing, the principal acquisition cost is the contract itself plus potential CBT and qualifying-offer consequences. In a trade, the acquiring club must also surrender prospects or other assets. In an extension, the club compares future team-specific wWAR over the extension years with the present value of guaranteed salary and the option value of existing club control. The same framework can therefore support multiple transaction types without redefining the underlying baseball metric.

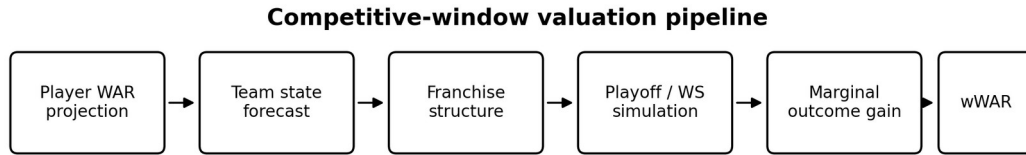
$$\text{NetAcquisitionValue}_{p,j} = \text{TAAV}_{p,j} - \text{ProspectPackageValue} - \text{Draft/PoolCosts}$$

This equation should not be interpreted as claiming that prospect packages or draft consequences are already perfectly valued in the current model. Rather, it identifies the proper architecture: wWAR estimates the destination-specific baseball benefit, while salaries, tax effects, prospects, draft picks, international bonus-pool penalties, and other assets enter as separate costs. A front office can plug its own internal valuations into those downstream terms.

7. Reproducibility and Generalization to Other Players

A useful player-value metric cannot remain a one-off Ohtani experiment. The final model is therefore organized as a frozen scoring engine. All learned team-strength parameters, franchise structural baselines, endpoint schedules, postseason calibration, uncertainty distributions, and fixed-reference curves are saved as model artifacts. Scoring a new player requires an annual WAR forecast vector and the player's current club; it does not require re-estimating the model.

The reproducible workflow has six steps: (1) generate or import annual projected WAR for the player; (2) identify the current club; (3) neutralize any player contribution already embedded in Hybrid current-state horizons; (4) insert the same projected WAR path into the actual team or all 30 candidate destinations; (5) simulate paired baseline and counterfactual universes with fixed model parameters; and (6) convert marginal championship and playoff effects through the published fixed reference to wWAR and pWAR. If the player is a free agent, the analyst can score every destination and compare the resulting distribution.



Hybrid short-run signal fades toward structural franchise strength; the same player is inserted into each destination under common random numbers.

Primary endpoint: expected World Series titles. Secondary endpoint: expected playoff appearances.

Figure 14. Reproducible scoring pipeline. Once model parameters are frozen, assigning wWAR to another player is a forward scoring problem: supply a WAR path and current team, then run the same counterfactual engine.

Reproducibility also requires explicit random-state control. The production and replication notebooks store simulation seeds, batch structure, input CSVs, and output tables. The final Ohtani result was separately challenged with 100,000 new universes generated from an independent seed sequence. That replication confirmed the sensitive NYY-LAD ordering as well as the much larger NYY-COL and LAD-COL contrasts. For future players, the same independent-seed check can be triggered when two destinations are close enough that ranking uncertainty matters.

The companion scoring utility distributed with this revision accepts a five-row projected-WAR CSV and a current-team code, then returns wWAR and pWAR for the current club or all 30 teams. This makes the proposed statistic operational rather than purely conceptual. The user can replace the Ohtani WAR path with any internally preferred projection system, allowing wWAR to sit downstream of ZiPS, Steamer, proprietary forecasts, or a team-specific projection model.

8. League-Wide Implications and Applications

8.1 Marginal utility and transaction asymmetry

One of the most consequential implications is that the same player can rationally have different reservation prices for different clubs. Figure 15 shows the relationship between a franchise's existing five-year championship window and the additional championship probability created by Ohtani. The relationship is strongly positive, but it is not identical across teams because division strength, seeding, postseason format, and future structural strength also matter.

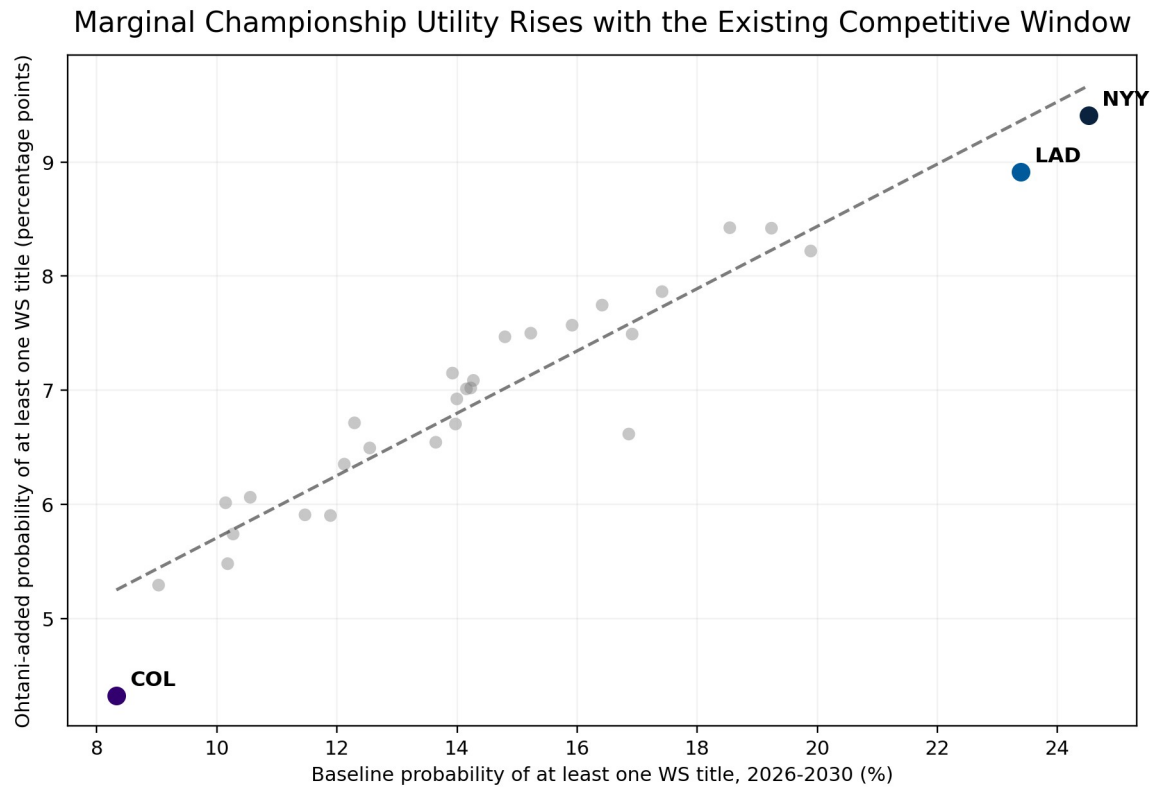


Figure 15. Marginal championship utility of the same player across MLB destinations. A player creates more championship value where the franchise is already positioned to convert added WAR into postseason success.

This asymmetry creates a natural economic basis for trades. A rebuilding club may correctly value an expensive veteran below the contender that acquires him, even when both clubs share the same projection of the player. The seller can exchange low local wWAR for prospects whose control years align with its future window, while the buyer converts the veteran's remaining WAR into higher immediate championship leverage. In this sense, wWAR formalizes part of the logic behind deadline buying and selling.

8.2 Roster construction and contract-window alignment

Because wWAR uses the annual timing of projected WAR, it can evaluate whether contract years align with a club's competitive window. A five-year contract whose production is front-loaded can be disproportionately valuable to a current contender, while the same total WAR delivered later may fit a team whose structural window is expected to strengthen. This creates an analytically cleaner way to compare short high-AAV contracts, long contracts, extensions, and players with different aging curves.

The distinction can also improve roster sequencing. Clubs can evaluate not only the highest raw-WAR upgrade, but the position and year in which an incremental win has the greatest marginal effect on playoffs or championships. pWAR and wWAR provide separate objective functions for teams whose near-term goal is simply reaching October versus teams already optimized for championship conversion.

8.3 Competitive balance and league policy

At league scale, wWAR offers a different lens on competitive balance. Payroll disparity matters not only because richer clubs can purchase more raw WAR, but because some franchises repeatedly operate in states where additional WAR has unusually high championship leverage. A league in which the same free-agent win systematically converts into more expected titles for a subset of franchises can exhibit a second layer of competitive advantage beyond the ability to pay for the player in the first place.

This does not imply that the current wWAR model should be used directly as a labor-policy formula. It does suggest measurable questions for future research: How concentrated is league-wide marginal championship

leverage? Do CBT rules reduce or reinforce that concentration? How does the distribution change under alternative playoff formats, salary thresholds, or revenue-sharing structures? Because the model is joint and counterfactual, those policy experiments are technically feasible without changing the underlying player metric.

8.4 Public analytics and communication

For public analysis, wWAR can complement rather than compete with WAR. Raw WAR remains the clean player-performance statistic. wWAR communicates the strategic value of those wins to a particular club. A transaction article, broadcast, or front-office dashboard could therefore show both numbers: projected WAR describes the player; wWAR describes the player-team fit. The contrast is especially useful when explaining why a deadline acquisition that looks expensive on a league-neutral basis can still be rational for a contender.

9. Limitations

- Championship outcomes are rare. Historical World Series probability validation has only a small number of observed champions, so championship calibration is inherently lower-power than regular-season or playoff validation.
- wWAR is prospective and model-based, not a causal estimate from randomized player movement. It evaluates counterfactual player insertion under the assumptions of the forecasting and simulation architecture.
- The player WAR path is an expected-value forecast. Player-specific injury, aging, role, and two-way usage uncertainty are not fully propagated as a joint distribution through every final counterfactual.
- Future franchise behavior is represented through structural historical persistence rather than explicit forecasts of every prospect, trade, free-agent signing, ownership change, front-office decision, or stadium effect.
- The standardized fixed reference is local to the tested WAR dose. Nonlinearity exists, although dose-response tests and fixed-reference stability checks indicate that the chosen +5 WAR scale is usable for the targeted application.
- The model values MLB baseball outcomes only. Marketing value, media value, attendance, merchandise, sponsorship, and player-specific commercial effects are excluded from wWAR and would need to be added downstream in a broader enterprise-value model.
- Economic results are particularly sensitive to the market \$/WAR assumption, CBT treatment, discount rate, contract structure, and transaction costs. TAAV should therefore be treated as scenario analysis rather than a single intrinsic dollar value.
- Prospect packages, draft-pick value, qualifying-offer penalties, and international bonus-pool effects are identified in the transaction framework but are not yet estimated by one unified public asset-pricing model.
- The Ohtani case is intentionally demanding because he is a two-way superstar. The scoring engine generalizes to other players, but a larger multi-player validation set is still needed before wWAR is treated as a universal public leaderboard.

10. Conclusion

Window-Adjusted WAR is a prospective extension of WAR that measures how the same projected baseball production converts into championships across different MLB franchises. The final model combines temporally validated player forecasts, a learned transition from current-team state to persistent franchise structure, joint five-year simulation, paired counterfactual player insertion, and a fixed league championship reference.

The final Ohtani case study provides a sharp demonstration. A fixed 19.81 projected WAR becomes 29.32 wWAR for the Yankees, 27.74 for the Dodgers, and 11.06 for the Rockies in the frozen 50K production model. An independent 100,000-universe seed replication confirms the Yankees-Dodgers ordering and the much larger elite-franchise versus Colorado gap. The result is not that WAR is wrong; it is that player production, championship leverage, and transaction value are different quantities.

The economic extension shows why that distinction matters to front offices. The same player contract can produce positive team-adjusted asset value for a contender and deeply negative value for another club because the

baseball benefit differs even when salary does not. The reproducibility framework also turns the model from a case study into a scoring system: any player with an annual WAR projection can be assigned wWAR on the player's current team or evaluated across all 30 destinations using the same frozen engine.

The broader research agenda is therefore clear: build multi-player wWAR leaderboards, integrate prospect and transaction-cost models, test trade-deadline decisions retrospectively, evaluate contract-window alignment, and use the simulator to study how MLB economic rules and playoff formats affect the distribution of marginal championship opportunity across the league.

References

- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324-345. <https://doi.org/10.2307/2334029>
- FanGraphs Library. Slowinski, P. (2012). WAR. <https://library.fangraphs.com/war/>
- Major League Baseball. Competitive Balance Tax. MLB Glossary. <https://www.mlb.com/glossary/transactions/competitive-balance-tax>
- Major League Baseball. (2023). Shohei Ohtani's Dodgers deal and deferred money, explained. <https://www.mlb.com/news/shohei-ohtani-dodgers-deal-deferrals-explained>
- Major League Baseball. (2026). MLB postseason format, explained. <https://www.mlb.com/news/mlb-playoff-format-faq>
- Major League Baseball Players Association. (2022). 2022-2026 Basic Agreement, Competitive Balance Tax provisions. <https://www.mlbplayers.com/>
- SABR / Lahman Baseball Database. (2025 release). Baseball data covering 1871-2025, originally created by Sean Lahman and maintained by SABR. <https://sabr.org/lahman-database/>
- Wharton Sports Analytics and Business Initiative. (2025). MLB payroll data, 02_mlb-payrolls.csv. https://github.com/WhartonSABI/lab-materials/blob/main/2025/labs/data/02_mlb-payrolls.csv

Appendix A. Selected Validation Statistics

Validation test	Metric	Result
Player forecast H1 vs persistence	MAE improvement = 0.204 WAR	Pass
Player forecast H2 vs persistence	MAE improvement = 0.272 WAR	Pass
Player forecast H3 vs persistence	MAE improvement = 0.326 WAR	Pass
Player forecast H4 vs persistence	MAE improvement = 0.361 WAR	Pass
Player forecast H5 vs persistence	MAE improvement = 0.384 WAR	Pass
Hybrid team strength H1 vs CurrentState	MAE improvement = 0.190 wins; 95% bootstrap interval 0.091 to 0.288	Pass
Hybrid H1 playoff probability vs naive	Brier skill = 13.4%; AUC = 0.725	Pass
H2 World Series endpoint ablation	Structural Brier 0.03173 vs Hybrid 0.03247; LOSO Hybrid weight = 0%	Use Structural
Cheap championship heuristic	Historical LOSO R2 = 0.778; forward Ohtani wWAR R2 = 0.842	Keep full simulator
Fixed WS reference stability	Future mean / 2021-2025 historical mean = 1.107	Pass
Independent NYY-LAD seed replication	Replication difference = +1.858 wWAR; 95% interval 0.673 to 3.044	Confirmed

Appendix B. Probability Accounting

The final production audit verifies that annual playoff probabilities sum to exactly 12 teams and World Series probabilities sum to exactly one champion in every simulated season, up to floating-point error on the order of 10^{-15} . This confirms that the league simulation preserves the joint outcome constraints rather than estimating each team independently.

Appendix C. Reproducibility and Any-Player Input Schema

The companion scorer expects an annual player projection table with two required columns: `forecast_season` and `projected_WAR`. A current-team code is supplied separately. The output contains raw projected WAR, marginal expected championships, marginal expected playoff appearances, wWAR, and pWAR for the requested destination or all 30 teams. No model refitting occurs during scoring.

Input	Example	Purpose
<code>forecast_season</code>	2026	Season in evaluation horizon
<code>projected_WAR</code>	4.2	Expected MLB WAR in that season
<code>current_team</code>	NY Yankees	Controls Hybrid baseline neutralization
<code>target_team</code>	NY Yankees or ALL	Club on which wWAR is assigned

For publication-quality replication, the analyst should also retain the model version, fixed-reference curve, random seed, simulation count, and input projection source. Close destination rankings should be confirmed with an independent seed sequence using the same frozen specification, as done for the Yankees-Dodgers comparison in this paper.