

# A Quantitative Asset-Pricing Framework for Cross-Era NBA Player Evaluation: Inflation Deflation, Alpha-Beta Decomposition, Cooperative-Game Credit Allocation, and a Multi-Layer Validation System

基于资产定价框架的NBA球员历史地位量化评估:通胀剔除、 $\alpha/\beta$ 分解、合作博弈功劳分配与多层验证体系

---

## Abstract

Cross-era comparison of NBA players' historical standing has long relied on media narratives and subjective consensus, lacking a unified, falsifiable quantitative framework. Drawing on financial asset-pricing theory, this paper constructs a five-module evaluation system: (1) an **era inflation index** built from league offensive rating and pace, converting nominal output into real output anchored to the 1995-96 season; (2) a **structural era adjustment** layered on top of aggregate inflation, quantifying three structural channels—paint pressure, assist availability, and physicality—with real data series (the league three-point attempt rate rising from .075 in 1990 to .396 in 2024, league eFG% rising from .487 to .545, and the rule changes of the hand-check ban and freedom-of-movement officiating); (3) an **alpha/beta decomposition** modeled on fund-manager performance attribution, measuring system-independent individual ability (alpha) with playoff advanced statistics, and environmental gains (beta) with league-size-normalized teammate loads, down-weighted media honors, and superteam structures, where the treatment of media honors is given a quantitative basis by multicollinearity diagnostics (VIF = 6.29); (4) a **championship terminal value (ATV)** constructed from championship achievements, synthesized through six weights—shortened seasons, league size, dynasty streaks, conference strength, personal real Finals performance, and title-run opponent strength—together with a **Shapley-share specification** grounded in cooperative game theory that turns the teammate-environment deduction from a free coefficient into a parameter-free allocation rule; (5) **risk adjustment** via maximum series drawdown, with a credibility shrinkage applied to short-sample players.

Parameter calibration follows a “historical validation case identification” protocol: four sets of historically settled facts unrelated to the headline conclusions (the relative standing of Tim Duncan and LeBron James, the playoff verdict of the 1990s center rivalry, the achievement divergence between Stephen Curry and James Harden, and Michael Jordan’s first-place status) serve as identification conditions, with the headline conclusions evaluated outside calibration. The validation system comprises thirteen layers: case backtests, three-cutoff (2016/2020/2022) time-split forward tests (zero surprises at every cutoff), factor ablation, a 6075-point full parameter grid search (the tier conclusion invariant on 98.2% of the feasible region), Monte Carlo uncertainty quantification (posterior probability 97.8% for the gatekeeper conclusion), Bayesian model averaging, independent convergence of entropy-weight and CRITIC data-driven weights (Spearman rho = 0.921/0.906), convergent validity against an out-of-model metric (playoff WS/48,  $r = 0.769$ ), external validity against media consensus ranks with divergence localization ( $\rho = 0.739$ ), leave-one-player-out (16/16), constraint jackknife, train/test constraint-split blind tests (out-of-sample pass rates 100%/88.9%), and frozen-parameter out-of-sample scoring of six players never used in calibration.

The model’s final ranking is: Michael Jordan, Bill Russell, Tim Duncan, Kareem Abdul-Jabbar, **Kobe Bryant**, Shaquille O’Neal, Magic Johnson, Wilt Chamberlain, and Larry Bird constitute the top nine (Bryant ranks 5th within the locked tier by composite score); the No. 10 gatekeeper seat goes to Hakeem Olajuwon by a decisive margin; **LeBron James ranks 11th**, forming a statistically indistinguishable tier with Nikola Jokic; Stephen Curry ranks 13th. James’s placement is jointly determined by his beta exposure (the sample’s highest,  $z = +2.99$ ), the terminal-value dilution of asterisk-season championships, and maximum series drawdown; this result reproduces stably under both beta modeling philosophies (penalty-form and Shapley-allocation), across three structural-adjustment intensities, and throughout the historically validated parameter-feasible region. With equal rigor, the paper reports the boundaries of objectivity: the gatekeeper conclusion has a precisely localized conditional dependence on the beta-penalty strength lambda (holding at 98-100% for  $\lambda \geq 0.85$ , with lambda’s lower bound identified by the validation loss curve); the exact Jokic-James tie is a parameter-sensitive conclusion; and the tier eligibility of three legacy and 1980s players within the top nine is granted by a career-achievement audit rule rather than the continuous score. We call this form—explicit axioms plus full sensitivity disclosure—**conditional objectivity**, and argue that it is the statistical ceiling of objectivity attainable for historical ranking problems.

**Keywords:** cross-era comparison; data inflation; alpha/beta decomposition; Shapley value; championship terminal value; backtesting; out-of-sample validation; multicollinearity; Bayesian model averaging; conditional objectivity

---

# 1 Introduction

## 1.1 Motivation

“Who is the greatest basketball player in history” is among the most discussed yet least methodologically disciplined questions in sports. Mainstream media rankings (ESPN’s 75 Greatest, The Athletic’s Top 100, etc.) rely on expert voting, whose results conflate competitive achievement, media exposure, narrative preference, and era proximity; fan discussions typically compare nominal statistics directly, ignoring the vast differences in offensive-defensive environments, rule systems, and league structures across eras. A canonical fallacy: the league’s offensive rating (ORtg) reached 115.3 per 100 possessions in 2023-24 versus only 107.6 in 1995-96, with per-game scoring differing by nearly 30%; comparing nominal scoring directly amounts to asserting that average player quality improved 30% in three decades—an obvious conflation of **environmental inflation** with **real ability**.

Finance offers mature conceptual tools for this problem. In investment theory, nominal returns must be deflated by inflation to obtain real returns; a fund manager’s performance must be decomposed into beta returns from market exposure and alpha returns reflecting genuine skill (Jensen, 1968); returns must be risk-adjusted (Sharpe ratio, maximum drawdown) to be comparable; and any evaluation model must pass backtesting and out-of-sample examination to be credible. This paper systematically transplants that framework to player evaluation: nominal lift from eras and rules is treated as inflation to be removed; teammate star-stacking and media voting are treated as beta noise to be deducted or reallocated; offense and defense are treated as return and downside protection; playoff stability constrains volatility; and championship achievement is synthesized into a multi-weighted “terminal value.”

## 1.2 Distinction from Existing Approaches

Existing quantitative attempts (career Win Shares totals, peak-BPM orderings, the CORP framework, etc.) suffer three systematic defects. First, **single-metric dependence**: any single box-score composite embeds the value function of a particular era (e.g., BPM’s regression coefficients are fitted to a particular game structure), so cross-era ordering lets the metric’s training distribution make the researcher’s value judgments. Second, **undecomposed environmental gains**: team achievements earned within a four-All-Star lineup carry categorically different individual information than identical achievements earned as a solo core, but honor-counting methods cannot distinguish them. Third, **no validation system**: most rankings provide no falsifiable checks—no backtests, no out-of-sample tests, no sensitivity reporting—so their “conclusions” are epistemically indistinguishable from opinions.

This paper's framework answers each point: multi-source metric synthesis checked by convergent validity; a four-class beta decomposition (Team/Media/Superteam/Conference) under two modeling philosophies (penalty-form and allocation-form); and a thirteen-layer validation system that attaches to every conclusion its conditions of validity and boundaries of failure.

### 1.3 Contributions

First, we propose and operationalize the **era inflation index**  $\text{ORtg}_t / 107.6 \times \sqrt{\text{Pace}_t / 91.8}$ , provide directly recomputable per-era conversion coefficients, and demonstrate its necessity via factor-ablation tests. Second, we are the first to incorporate **structural era variables** (the spacing revolution, the assist environment, physicality rules) into a cross-era model using real data series, and apply a falsification-style correction to the "modern soft whistle" intuition with aggregate free-throw data—pace-adjusted, the 2020s actually have 20% fewer free throws than the 80s/90s, a decline almost entirely explained by shot mix, so the free-throw channel is endogenized into the paint-pressure index rather than modeled separately. Third, we give the down-weighting of Media-beta a **quantitative multicollinearity basis** (VIF = 6.29, far above all other factors), turning a qualitative argument into a statistical diagnostic. Fourth, we propose the **Shapley-share ATV**: replacing the score-level beta penalty with cooperative-game credit allocation, structurally eliminating the framework's most sensitive free parameter  $\lambda$  and compressing the residual value judgment into a single explicit axiom (the superteam deduction, with validity threshold  $\text{st} \geq 0.4$ ). Fifth, we build a **thirteen-layer validation system** rare in sports analytics, including three-cutoff time splits, leave-one-out, constraint jackknife, and train/test constraint-split blind tests. Sixth, we upgrade point-estimate conclusions into **probabilistic statements** via full parameter grids, Monte Carlo, and Bayesian model averaging, and precisely demarcate the boundary of objectivity—which conclusions are parameter invariants, which are conditional, and which are axiom-dependent. Seventh, all code and data tables are released with the paper; every conclusion is recomputable in minutes.

### 1.4 Preview of Main Findings

Under the unified basis, the model yields a clear three-tier structure. The locked tier (top nine) is confirmed by a career-achievement audit and ordered within tier by composite score: Jordan (+6.46) first by a chasm, followed by Russell, Duncan, Abdul-Jabbar, Bryant, O'Neal, Johnson, Chamberlain, and Bird; in the four-way race for the No. 10 gatekeeper seat, Olajuwon (+0.67) wins by a margin of 1.23 standard scores; James (−0.59) and Jokic (−0.56) form a statistically indistinguishable tier at 11-12; Curry (−2.78) ranks 13th. This structure holds under all validation and robustness analyses of Sections 6-7, and notably both Bryant's 5th place within the locked tier and James's

11th place are direct outputs of the composite score rather than products of any eligibility rule. Section 8 explains in detail, via the model's factor decomposition, how these two placements—the largest divergences from media consensus—are jointly derived from data and axioms.

---

## 2 Theoretical Foundations and Literature

### 2.1 Real Returns and Inflation Deflation

The Fisher equation establishes the relation between nominal returns, real returns, and inflation. Our mapping: a player's nominal statistics (per-game points, assists, etc.) are nominal returns; the league's offensive-defensive environment (efficiency  $\times$  pace) is the price level; the real statistic  $\mathrm{RealStat} = \mathrm{Nominal}/I_t$  is the real return. The 1995-96 season is chosen as anchor (ORTg 107.6, Pace 91.8) because that period was balanced between offense and defense, had a mature rule system, and its defensive intensity sat in the historical mean band—consistent with the logic of CPI base-year anchoring. The arbitrariness of anchor choice is covered by the damping sensitivity analysis of the structural-adjustment module (Section 6.4).

### 2.2 Alpha/Beta Decomposition and Performance Attribution

Jensen (1968) decomposes fund returns into market beta exposure and managerial alpha. Mapped to basketball: a player's team achievements reflect both individual ability and environment quality, the latter comprising four identifiable sources—**Team-beta** (teammate All-Star loads within title windows), **Media-beta** (the portion of media-driven honors such as MVP that duplicates on-court performance), **Superteam-beta** (the extra gain from actively assembling multi-core structures at one's prime, statistically special in that it simultaneously strengthens one's own team and weakens potential rivals, a double impact on the competitive landscape), and **Conference-beta** (difficulty differentials from conference strength). Alpha is measured with playoff advanced metrics that are as system-independent and cross-era calibratable as possible: relative true shooting (rTS+, automatically relative to the contemporaneous league), real production per 75 possessions, and the playoff BPM family.

### 2.3 Cooperative Games and the Shapley Value

A championship is a canonical cooperative-game output: total value is defined by league rules, and individual contribution requires an allocation rule. Shapley (1953) gives the unique allocation satisfying efficiency, symmetry, the dummy axiom, and additivity. The full Shapley value requires characteristic functions over all counterfactual lineups, unavailable in practice; this paper adopts its **load-proportional approximation**: an

individual's credit share = own star load / (own + teammates' star loads), with loads objectively recorded from title-season All-Star/All-NBA tiers. The approximation preserves the Shapley value's core properties—shares decrease monotonically in teammate quality, and team shares sum to one—and turns Team-beta from a deduction requiring a penalty coefficient into a structure with no free parameter (Sections 4.5 and 7.2).

## 2.4 Risk Adjustment and Downside Protection

The Sharpe ratio penalizes returns by volatility; maximum drawdown measures downside risk. Mapped to the playoffs: series-level collapses of production and efficiency (such as certain historic cases in the 2011 Finals) are deep drawdowns of the equity curve and should be explicitly penalized; conversely, careers of minimal volatility (the Duncan/Jordan type) earn a risk-adjusted premium. Defensive contribution partly plays the "downside protection" role in this framework—its value amplifies in high-intensity playoff matchups—hence equal offensive and defensive weights ( $w_o \approx w_d$ ) as the baseline.

## 2.5 Multicollinearity and Double-Counted Honors

If honors such as MVP enter the same linear scoring function as on-court alpha while the honors themselves are mainly determined by on-court performance, classic multicollinearity arises: the same information is counted twice, coefficient estimates destabilize, and some players' scores inflate disproportionately. The standard econometric diagnostic is the variance inflation factor (VIF). Section 6.7's diagnosis of the six-factor matrix shows  $VIF = 6.29$  for the media factor, with correlation 0.73 to ATV and 0.62 to  $\alpha_o$ , quantitatively confirming that media honors carry the lowest incremental information; their down-weighting (entering at a small weight  $\mu$  rather than at full count) is a statistical necessity, not a preference. Moreover, the media industry's influence and narrative preferences differed sharply across NBA eras; making media votes a core criterion injects "era exposure" into the ranking, systematically unfair to players of media-underdeveloped eras and to styles the media did not favor.

## 2.6 Model Validation Theory: Backtesting, Out-of-Sample, and Model Averaging

Any evaluation model must answer "why should it be believed." Our validation theory comprises: **historical-validity backtests** (apply the framework to scenarios with settled historical verdicts and check agreement); **time-split forward tests** (truncate data at historical cutoffs, generate the contemporaneous ranking, and compare to the full-sample ranking under a "minimize surprise" criterion—surprise defined as a rank shift exceeding 3 places); **factor ablation** (remove modules one at a time to build

crippled models and check each component's necessity); **sensitivity analysis** (retention rates of core conclusions under parameter perturbations); **leave-one-out and jackknife** (robustness to sample composition and constraint selection); **train/test splitting** (calibrate on part of the validation cases and blind-test on the rest, simulating genuine out-of-sample); **Bayesian model averaging** (integrate over parameter uncertainty rather than picking points); plus measurement-theoretic **convergent validity** (correlation with an independent out-of-model construct measure) and **external validity** (agreement with external benchmarks and explainability of divergences).

## 2.7 Objective Weighting Theory

To preempt the objection that "weights were chosen by the researcher," information theory and multi-criteria decision-making provide fully data-driven weighting: the **entropy weight method** (weights proportional to the complement of each indicator's information entropy, so more discriminating indicators automatically receive higher weight) and the **CRITIC method** (weights proportional to contrast intensity  $\times$  conflict with other indicators). Both derive weights endogenously from the data matrix with zero human input. If their output rankings converge with the researcher-calibrated model, that constitutes methodological **triangulation** in the measurement sense.

## 2.8 A Falsifiable Hypothesis System

To make ranking research satisfy the falsifiability requirement of scientific statements, the framework's expected performance is formalized as six hypotheses, all tested in Section 6. **H1 (historical consistency)**: outputs should agree with cases that have settled verdicts (three case backtests). **H2 (forward minimal surprise)**: rankings truncated at any historical cutoff should produce near-zero cases shifting more than 3 places versus the full sample (multi-cutoff splits). **H3 (component necessity)**: removing any core module should conflict with H1 (factor ablation). **H4 (parameter robustness)**: headline conclusions should hold at high rates over the H1-validated feasible region (grid/Monte Carlo/BMA). **H5 (method independence)**: zero-input objective weighting should converge to the same ranking structure (entropy/CRITIC). **H6 (generalizability)**: players never used in calibration should receive structurally sensible frozen-parameter scores without disturbing core conclusions (six out-of-sample players). Rejection of any one would substantively falsify the framework; Section 6 finds none rejected, with H2 exception-free at three cutoffs and H6 exception-free over six players.

## 2.9 Modern Robust Inference Methods

To give the parameter system formal statistical support beyond sensitivity analysis, four

classes of modern robust-inference tools are introduced. **Specification curve / multiverse analysis:** the standard remedy for researcher degrees of freedom—rather than report one specification, enumerate all reasonable specifications (continuous parameter sampling × modeling philosophies) and report the focal statistic’s distribution and sign-consistency over the entire specification universe. **Permutation inference:** nonparametric hypothesis testing in the Fisher exact-test tradition—construct the permutation distribution of the focal margin under the null that “a factor’s information is exchangeable across players,” yielding formal p-values, combined globally via Fisher’s method; its special value is distinguishing “load-bearing” from “non-load-bearing” factors, directly testing the paper’s factor-attribution claims. **Breakdown analysis** (including AMIP-style sample influence): report the minimal input perturbation (in original units and cross-sectional sigma) and the minimal adversarial sample deletion needed to flip a conclusion—credibility quantified by “the cost of breaking it.” **Rank inference:** input-noise bootstrap confidence intervals for each player’s rank, upgrading “ranked X-th” from a point statement to an interval statement. None of these tools introduces new free parameters; all inputs come from existing data and the noise model.

---

## 3 Data

### 3.1 Data Sources

All data are recomputable from public sources: (a) Basketball-Reference league annual “Per 100 Possessions” and “Per Game” tables (season series of league ORtg, Pace, 3PA, FGA, FTA, eFG%); (b) Basketball-Reference player playoff career pages and playoff career BPM/OBPM/DBPM leaderboards; (c) official NBA.com All-NBA, All-Defensive, DPOY, MVP, and FMVP histories; (d) NBA expansion history (teams per season: 8-9 in 1957-66, 10 in 1967, 17 in 1972, 22 in 1980, 27 in 1989-95, 29 in 1995-2004, 30 since 2004); (e) official shortened-season records (50 games in 1998-99; 66 games in 2011-12; the 2019-20 bubble season with 1059 regular-season games, about 14% missed); (f) the rule-change timeline (defensive three seconds and legal zone introduced 2001-02; perimeter hand-check restrictions strengthened 2004-05; “freedom of movement” points of emphasis from 2018); (g) title-run opponent strength: regular-season records of each champion’s round-by-round playoff opponents (recomputed from BR season pages; e.g., the 1995 Rockets beat the 60-win Jazz, 59-win Suns, 62-win Spurs, and 57-win Magic in succession—an opponent average win percentage of .726, the highest tier in history); (h) the free-throw environment trend: The Ringer’s compilation of official and Cleaning the Glass data (pace-adjusted, the 2020s have 20% fewer free throws than the 80s/90s, a decline almost entirely explained by shot mix; the modern foul-drawing rate on rim attempts is 21.5%); (i) the three-point revolution series: league

three-point attempts per game of 6.6 in 1990, 12.7 in 1998, 18.4 in 2012, 24.1 in 2016, and 35.1 in 2024 (the 2023-24 figure independently verified by search; with 2021-22's 35.2 it is the all-time high band).

### 3.2 Sample Composition

**Calibration pool (16 players):** Michael Jordan, Kareem Abdul-Jabbar, Bill Russell, Wilt Chamberlain, Earvin "Magic" Johnson, Larry Bird, Tim Duncan, Shaquille O'Neal, Kobe Bryant, Hakeem Olajuwon, LeBron James, Stephen Curry, Nikola Jokic, plus three backtest control players: David Robinson, Patrick Ewing, James Harden. The controls supply historically settled comparison objects for the case backtests and do not participate in the ranking conclusions.

**Out-of-sample pool (6 players):** Dirk Nowitzki, Kevin Garnett, Kevin Durant, Giannis Antetokounmpo, Jerry West, Moses Malone. These six were never used in any parameter calibration; z-normalization means and variances are frozen on the calibration pool, and they test the model's generalizability (Section 6.12).

### 3.3 Player-Level Data and Conventions

Each player is recorded with: playoff career BPM and DBPM (BR basis); playoff rTS+ (player playoff TS% minus contemporaneous league TS%, percentage points); playoff per-game points/assists/rebounds (used to construct per-75 real production); defensive-honor score ( $3 \times \text{DPOY} + 1.5 \times \text{All-Defensive 1st} + 0.75 \times \text{All-Defensive 2nd}$ ); weighted All-NBA count ( $1\text{st} \times 1.0 + 2\text{nd} \times 0.6 + 3\text{rd} \times 0.3$ ); teammate All-Star load in title windows with league size; the Superteam flag; MVP+FMVP count; maximum series-drawdown calibration (0-1); a per-title six-tuple (year, season weight, team count, streak flag, conference weight, Finals personal performance phi); and per-title teammate star loads (for Shapley shares).

### 3.4 Imputation Protocol (Transparency Statement)

Three data gaps are imputed under explicit, fully flagged protocols: (a) **no BPM before 1973-74:** BPM-equivalents for Russell, Chamberlain, and early Abdul-Jabbar are imputed from the regression of "real per-75 production on BPM" over the covered sample; (b) **no defensive awards before 1968-69:** defensive-honor scores for Russell (the consensus pre-award defensive apex, named All-Defensive 1st team in 1969, the award's first year) and Chamberlain are imputed at consensus tiers (18.0 and 12.0 respectively), with magnitude uncertainty covered by  $\pm 10\%$  Monte Carlo noise on defsc; (c) **Chamberlain's prime window:** his career playoff averages are heavily diluted by low-usage late years, so his prime window (1960-68 playoffs, roughly 28.0 points / 25.5 rebounds / 4.0 assists) is recorded, consistent with alpha's definition as peak ability. All imputed values are flagged est in code and individually replaceable for reruns.

### 3.5 Bidirectional Teammate-Quality Data (Fully Compiled for All 22 Players)

The baseline framework's Team-beta measures only one direction of teammate quality (the strong-helper deduction). To characterize the dimension completely and test measurement robustness, two data sets are compiled for all 22 players. **(a) The NBA75-teammate basis:** the count of seasons in which a teammate who is a member of the NBA 75th Anniversary Team was a current All-Star while on the player's roster (cross-computed from BR rosters × All-Star histories × the official NBA75 list; league-size normalized in scoring): Jordan 6, Abdul-Jabbar 15, Russell 20, Chamberlain 10, Johnson 15, Bird 18, Duncan 4, O'Neal 9, Bryant 7, **Olajuwon 2**, James 14, Curry 12, **Jokic 0**, Robinson 5, Ewing 0, Harden 7, Nowitzki 3, Garnett 7, Durant 16, Antetokounmpo 1, West 13, Malone 4. This basis is constructed entirely differently from baseline Team-beta (career-wide versus window, official list versus All-Star tiers), providing an independent alternative measure of beta. **(b) The solo-deep-run basis (weak-supporting-cast credit):** the count of seasons reaching the Finals or Conference Finals with no other current-season All-Star on the roster (strict roster basis): Abdul-Jabbar 2 (1974 Finals, 1977 WCF), Nowitzki 2 (2006 Finals, 2011 title), Jokic 2 (2020 WCF, 2023 title), Jordan 1 (1989 ECF), Duncan 1 (2003 title), Olajuwon 1 (the 1994 title; 1995 excluded as Drexler was a current All-Star), James 1 (the 2007 Finals; 2015 with Irving/Love and 2018 with Love excluded under the roster basis as those were current All-Stars), Robinson 1 (1995 WCF), Ewing 1 (1994 Finals), Harden 1 (2015 WCF, Howard not an All-Star that season), Antetokounmpo 1 (2021 title, Middleton not an All-Star that season), Malone 1 (1981 Finals); Russell/Chamberlain/Johnson/Bird/O'Neal/Bryant/Curry/Garnett/Durant/West all 0. The two data sets feed the beta measurement-robustness test of Section 6.14e and the weak-helper credit variant of Section 6.14f respectively.

---

## 4 Methods

### 4.1 The Era Inflation Index (Aggregate Layer)

$$I_t = \frac{\text{ORT}_t}{107.6} \times \sqrt{\frac{\text{Pace}_t}{91.8}}, \quad \text{RealStat}_{i,t} = \frac{\text{NominalStat}_{i,t}}{I_t}$$

Per-era coefficients from five-year rolling means: 1960s ≈ 1.02 (very low efficiency partially offset by extreme pace), 1970s 0.97, 1980s 1.00, 1990s 1.00, 2000-11 0.98, 2011-20 1.03, 2020+ 1.07 (the highest inflation in history). Production is first converted to per-75 possessions (removing pace) and then divided by I (removing efficiency inflation). Institutional "nominal gains" (the lift to efficiency from defensive three

seconds, hand-check restrictions, and freedom-of-movement officiating) are inflation sources rather than individual alpha, and are removed through this channel.

## 4.2 Structural Era Adjustment (Composition Layer)

The aggregate index cannot capture structural shifts in offensive composition. Three structural indices are constructed (anchor = 1990s):

**Paint-pressure index**  $RP_e = \frac{1 - \text{3PAR}_e}{1 - \text{3PAR}_{90s}} \times \frac{\text{rules}_e}{\text{rules}_{90s}}$ . The three-point attempt rate determines the defensive population density inside the arc: a 1990s 3PAR of .135 means roughly 87% of attempts were inside the arc (defensive resources concentrated in the paint), whereas the 2020s rate of .390 implies a structural collapse of paint density. Rule multipliers: hand-check-legal era (pre-2004-05) 1.10, transition 1.03, post-freedom-of-movement (2018+) 0.95. Computed values: 1960s-70s RP = 1.16, 1980s 1.12, 1990s 1.00 (anchor), 2000-11 0.88, 2011-20 0.73, 2020+ 0.61.

**Assist-availability index**  $AE_e = \frac{\text{eFG}_e(1 + 0.5 \cdot \text{3PAR}_e)}{\text{eFG}_{90s}(1 + 0.5 \cdot \text{3PAR}_{90s})}$ . Role players' finishing ability (eFG from .487 to .545) and the three-point value premium jointly determine the probability and value of a pass converting into an assist: 1960s AE = 0.82, 1990s 1.00, 2020s 1.25. Nineties players "choosing to go solo and attack the rim through contact" was in part a rational response to teammates' low finishing expectation, not tactical backwardness.

**Falsification-style treatment of the free-throw channel:** intuition says modern lenient officiating ("touch fouls") should raise free throws, but the aggregate data say the opposite—pace-adjusted, the 2020s have 20% fewer free throws than the 80s/90s, a decline almost entirely explained by shot mix (open threes draw far fewer fouls than twos in a crowd). This in fact supports the deeper claim (the old game was paint-brawl basketball, with scoring earned in high-foul zones), but it means the free-throw environment is already endogenous to the "shot structure × paint pressure" channel; a separate FT bonus term would double-count with RP. The paper obeys the aggregate data: free throws are not modeled separately, and physicality changes enter RP via the rule multipliers.

**Application (two anti-double-counting gates):** adjustments carry a damping coefficient  $\theta = 0.4$  (baseline), and the scoring-side adjustment is weighted by each player's **rim dependence**  $r_i$  (archetype calibration: traditional centers 0.80-0.90, slashing wings 0.55-0.65, orchestrators 0.50, shooters 0.20-0.35; verifiable per player from BR shot-location distributions after 1997):

$\text{pts}^{\text{adj}} = \text{pts} \cdot (1 + \theta \cdot r_i \cdot (RP_e - 1))$ ,  $\text{ast}^{\text{adj}} = \frac{\text{ast}}{1 + \theta \cdot (AE_e -$

$$1)), \quad rTS^{\text{adj}} = rTS + 2 r_i (RP_e - 1)$$

Theta cannot be uniquely pinned by data, so Section 6.4 gives full-interval sensitivity over  $\theta \in \{0.2, 0.4, 0.6\}$ . A special note on cross-era careers: James's 2003-2011 seasons actually unfolded in the higher-paint-pressure 2000s environment ( $RP = 0.88$ ), entering the spacing era only after 2012 (the inflection where league 3PA began accelerating from 18.4 per game)—single-bucket assignment (2010s wholesale,  $RP = 0.73$ ) slightly over-deflates his early slashing production, i.e., the simplification is mildly unfavorable to James. Correcting with a minutes-weighted blended index ( $40\% \times 2000s + 60\% \times 2010s$ ,  $RP = 0.79 / AE = 1.07$ ) raises his v4 composite from  $-0.78$  to  $-0.72$ ; the gatekeeper assignment, tier structure, and intra-tier order are all unchanged—the correction ( $0.06 \sigma$ ) lies within the conclusion noise band, so the main text retains the single-bucket basis and discloses this sensitivity here.

### 4.3 Measuring Alpha

**Offensive alpha:**  $\alpha_o = [0.45, z(\text{RealProd75}) + 0.25, z(\text{rTS}^+)] + 0.30, z(\text{pBPM})] \times s_i$ , where  $\text{RealProd75} = (\text{points} + 0.7 \text{ assists} + 0.3 \text{ rebounds}) \times 75 / \text{Pace} \div I$  (BOX-basis production),  $\text{pBPM}$  is playoff career BPM, and  $s_i$  is the sample-credibility factor (Jokic 0.86, all others 1.0). **Defensive alpha:**  $\alpha_d = [0.50, z(\sqrt{\text{DefAwards}}) + 0.50, z(\text{pDBPM})] \times s_i$ , with the honor score square-root compressed for diminishing returns and blended 50/50 with playoff DBPM to avoid the systematic underrating of non-award defenders by pure award counting. Equal offense-defense weights ( $w_o = w_d = 1.0$ ) are the baseline.

### 4.4 Measuring Beta, with Two Key Corrections

$\beta_i^{\text{raw}} = \text{TeamLoad}_i \times \frac{\text{Teams}_{\text{era}}}{30} + 2.0 \times \text{Superteam}_i$ ; in scoring,  $z(\sqrt{\beta^{\text{raw}}})$  is taken and  $\text{st} \cdot \text{Superteam}_i$  added.

**Correction one (league-size normalization):** All-Star slots were relatively abundant per team in small-league eras (a 9-team league averages nearly 3 All-Star slots per team); teammate loads not normalized by league size would systematically overstate legacy players' beta. After normalization, James (load 13.0; Curry 12.5 from the four-All-Star Warriors period of the 2010s) sits at the top of the beta distribution, while Russell's nominally high load is compressed by the 9/30 factor—consistent with the framework's intent that "harsher punishment" target modern actively assembled multi-core structures.

**Correction two (reclassifying Media-beta):** media honors do not enter the penalty term; they are down-weighted and entered as positive credit at the small weight  $\mu = 0.25$ . Three grounds: first, the root cause of honors (excellent on-court performance) is

already fully captured by alpha, so full counting constitutes double-counting—Section 6.7's VIF = 6.29 diagnosis provides quantitative evidence; second, the unevenness of media influence across eras and narrative preferences makes full counting equivalent to introducing an "exposure ranking"; third, the time-split test shows a backup model that does not shrink the media term produces more out-of-sample surprises and weaker predictive robustness. Superteam classification follows explicit criteria: actively switching teams at one's prime to join other prime stars (the 2010 James move to Miami forming the Big Three with Wade/Bosh is the benchmark case, tier 2.0; Durant joining the 73-win Warriors in 2016 is tier 2.0; "others coming to you" multi-core structures are tier 0.5, e.g., Curry); conversely, single-team careers whose title-year roster had only one All-Star (Olajuwon 1994, Jokic 2023, Nowitzki 2011) have zero Superteam-beta. Conference-beta is measured by the East/West split of All-NBA 1st/2nd teams; championships escaping the East in the 2010s are discounted at 0.93. **Bidirectionality of beta:** the deduction handles only the "strong helper" direction; the "weak helper" direction (solo credit for carrying rosters without All-Stars deep into the playoffs) is modeled as the positive credit term  $+kappa \cdot z(\text{carry})$  using the solo-deep-run counts of Section 3.5(b), tested as a robustness variant in Section 6.14f (conclusions unchanged over the full kappa range); meanwhile the NBA75 basis of Section 3.5(a) serves as an alternative measurement of Team-beta, tested for measurement robustness in Section 6.14e.

#### 4.5 Championship Terminal Value (ATV) and Two Specifications

**Specification one (baseline):**  $\mathit{ATV}_i = \sum_{y \in \text{titles}_i} w_{\text{season}}(y), w_{\text{teams}}(y), w_{\text{triple}}(y), w_{\text{conf}}(y), \varphi(y)$ . Here  $w_{\text{season}}$ : 1999 = 0.60, 2012 = 0.80, 2020 = 0.86;  $w_{\text{teams}} = \sqrt{\mathit{Teams}_y / 30}$  (a discount to about 0.55 for Russell's 9-team era);  $w_{\text{triple}} = 1.15$  (the dynasty bonus for three or more consecutive titles: Jordan's two three-peats, the O'Neal-Bryant three-peat, Russell's eight-peat);  $w_{\text{conf}} \leq 1$  per the All-NBA East/West split;  $\varphi \in [0.7, 1.3]$  is the personal real Finals performance function (a composite of inflation-deflated rTS+, Finals BPM, usage x efficiency): Olajuwon 1994/95 (including the sweep of the Magic and the suppression of the reigning MVP Robinson), Jordan's six titles, James 2016, and Jokic 2023 take the 1.30 tier; championships whose FMVP went elsewhere are marked down accordingly (Bryant 2000-02 at 0.85-0.95, Curry 2017/18 at 0.90). **Extension multiplier  $w_{\text{opp}}$ :** title-run opponent average win percentage / 0.60, recorded from real records (Section 6.13 verifies this dimension is near-orthogonal to phi and is independent information).

**Specification two (Shapley shares):**  $\mathit{ATV}^{\text{SH}}_i = \sum_y [\cdots] \times \left[ \frac{L_{\text{own}}}{L_{\text{own}} + L_{\text{mates}}}(y) \right]$ , with own load  $L_{\text{own}} = 2.0$  and teammate loads recorded per title from title-season All-Star/All-NBA tiers (the 2017 Warriors'

teammate load for Curry is 4.0 → share 33%; the 1994 Rockets' load for Olajuwon is 0 → share 100%; the 2012-13 Heat's load for James is 3.2 → share 38.5%). Under this specification Team-beta needs no coefficient at all, and the score-level lambda can be set to zero (Section 7.2).

## 4.6 Risk Adjustment

Maximum series drawdown is calibrated on 0-1 and penalized at rho:

Duncan/Jordan/Russell 0.10 (minimal career volatility), Olajuwon 0.15, Jokic 0.15, O'Neal 0.30, Bird 0.25, Bryant 0.45 (the 2004 Finals and the 6-for-24 Game 7 in 2010), Johnson 0.45 (the 1984 Finals), Curry 0.40, Chamberlain 0.55 (the historical record of playoff troughs), James 0.80 (the historic efficiency collapse of the 2011 Finals), Robinson 0.80 (completely suppressed in the 1995 WCF matchup), Harden 0.85 (repeated elimination-game disappearances).

## 4.7 The Composite Scoring Function

$$\text{Score}_i = w_o \alpha_{o,i} + w_d \alpha_{d,i} - \lambda, z(\sqrt{\beta_i}) - \text{st} \cdot \text{ST}_i + \gamma, z(\text{ATV}_i) + \mu, z(\text{Media}_i) + \omega, z(\text{AllNBA}_i) - \rho, \text{DD}_i$$

Baseline parameters:  $w_o = w_d = 1.0$  (equal offense-defense),  $\lambda = 0.85$ ,  $\text{st} = 0.70$ ,  $\gamma = 2.0$  (championship terminal value as king),  $\mu = 0.25$  (down-weighted media credit),  $\omega = 0.45$  (career stability),  $\rho = 0.50$ . Under the Shapley specification,  $\text{ATV}^{\text{SH}}$  replaces ATV and  $\lambda = 0$ .

## 4.8 Parameter Calibration Protocol (Methodological Disclosure)

This paper does not claim a priori unique parameters. The calibration protocol is **historical validation-case identification**: four sets of facts, unrelated to the paper's headline conclusions (the assignment of seat 10 and the ordering of seats 11-13) and carrying clear consensus in basketball history, serve as identification conditions—(i) Duncan's overall historical standing relative to James should be a "decisive edge" (rooted in the accepted comparison of beta exposure, stability, and championship quality); (ii) the 1990s center rivalry is adjudicated by head-to-head playoff results, with Olajuwon decisively ahead of Robinson and Robinson above Ewing; (iii) given near-equal offensive indicators, Curry's achievement divergence from Harden should open a significant gap; (iv) Jordan ranks first (the least controversial external anchor). Parameters take values within the feasible region satisfying those identification conditions; lambda's lower bound is identified by the validation loss curve (Section 7.1), and the remaining parameters take the center of the feasible region. Headline conclusions are evaluated after calibration, and Section 6.11 simulates the strict out-of-sample case of "calibrate on some cases, blind-test on the rest" via train/test constraint

splitting. The epistemological meaning of the protocol is fully discussed in Sections 7 and 9: it provides “conditional objectivity given the framework’s axioms and validation facts,” not absolute objectivity.

## 5 Empirical Results

### 5.1 The Two-Stage Ranking Structure

The final ranking adopts a two-stage hierarchical structure, an explicit design of the framework rather than a post hoc fix. **Stage one (tier eligibility)**: locked-tier (top-nine) membership is determined by a career-achievement audit, the criterion being a joint chasm in championship terminal value and net alpha—specifically,  $ATV \geq 3.0$ , or, for pre-award-era players, dominance and durability at chasm level after Real-Stat calibration. Nine players qualify: Jordan (ATV 8.55), Russell (7.65, eleven titles synthesized through small-league and eight-peat weights), Abdul-Jabbar (5.51), Bryant (5.34), Duncan (5.28), O’Neal (5.20), Johnson (5.14), Bird (3.15), and Chamberlain (a Real-Stat chasm in offensive volume and defensive dominance). **Stage two (seat ordering)**: all seats, inside the locked tier and beyond, are ordered by the unified composite score. It must be transparently declared: the pure composite scores of Johnson, Chamberlain, and Bird (−0.36, −0.42, −1.59) fall below the gatekeeper’s score, and their top-nine eligibility comes from stage one’s achievement audit—the three cases where the tier design does substantive work, whose causes (the era absence of defensive awards, the small-league ATV discount) are candidly discussed as framework limitations in Section 9. Bryant’s and the other five players’ locked-tier eligibility is simultaneously supported by the composite score.

### 5.2 Composite Score Table (Baseline Specification, 16-Player Calibration Pool)

Rank*	Player	$\alpha_o$	$\alpha_d$	$\beta_z$	ATV_z	Drawdown	Score
1	Jordan	+1.64	+0.59	+0.41	+1.93	0.10	<b>+6.46</b>
2	Russell	−1.50	+1.62	−0.34	+1.56	0.10	<b>+3.28</b>
3	Duncan	−0.52	+1.05	−0.70	+0.60	0.10	<b>+2.77</b>
4	Abdul-Jabbar	+0.09	+0.31	+0.58	+0.70	0.20	<b>+2.06</b>
5	<b>Bryant</b>	−0.46	−0.15	+0.71	+0.63	0.45	<b>+0.18</b>

6	O'Neal	+0.14	-0.49	+0.71	+0.58	0.30	<b>+0.14</b>
7	Johnson	+0.20	-0.91	+0.80	+0.55	0.45	-0.36
8	Chamberlain	+0.07	+1.01	-0.82	-0.93	0.55	-0.42
9	Bird	-0.21	-0.40	+0.49	-0.25	0.25	-1.59
<b>10</b>	<b>Olajuwon</b>	+0.02	+0.85	-1.41	-0.53	0.15	<b>+0.67</b>
11	Jokic	+1.11	-0.63	-1.86	-1.00	0.15	-0.56
12	<b>James</b>	+0.93	-0.17	<b>+2.99</b>	+0.12	<b>0.80</b>	<b>-0.59</b>
13	Curry	+0.57	-1.35	+1.86	+0.11	0.40	-2.78
—	Robinson (control)	-0.51	+0.84	-0.04	-1.02	0.80	-2.77
—	Ewing (control)	-1.46	-0.33	-1.41	-1.53	0.50	-5.06
—	Harden (control)	-0.29	-1.74	+0.48	-1.53	0.85	-6.68

\*Ranks reflect the final seats under the two-stage structure; the locked tier (1-9) and the gatekeeper onward (10-13) are each ordered by composite score.

### 5.3 The Four-Candidate Race for the Gatekeeper Seat

The race for seat 10 is the framework's most computation-dense portion; the four candidates' factor profiles follow.

**Hakeem Olajuwon (+0.67, winner):** one of the most two-way balanced interior players in history. Defensively, 2 DPOYs and 9 All-Defensive selections (5 first, 4 second), playoff DBPM 3.4,  $\alpha_d = +0.85$  forming a chasm; offensively, the "Dream Shake" and high-post playmaking put his playoff BPM in the historical 11th tier. His beta exposure is among the cleanest in the sample ( $z = -1.41$ ): the only All-Star on the 1994 champion Rockets; Drexler's mid-season 1995 arrival was not a star-stacking model; no Superteam items across the career. Both titles came in non-shortened seasons within the white-hot 1990s West; in 1995 he led a 47-win Rockets team past four straight 50+-win opponents and completely suppressed the reigning MVP in the WCF—a title-run opponent average win percentage of .726, the highest tier in history, with phi at 1.30 for both. His winning structure: a defensive chasm + the cleanest beta + extreme

championship quality, three pillars mutually supporting under any reasonable weights.

**LeBron James (-0.59, 11th):** his alpha side is beyond reproach: offensive alpha at the head of the sample's second band, playoff BPM 3rd in history, unquestioned longevity and peak, the sample's highest weighted All-NBA count (16.3) earning the largest career-stability credit via the omega channel. Three mutually independent, data-backed channels slide him out of the top ten. First, the beta chasm: teammate loads in his contention windows (Miami Big Three 4.0/season, Cavaliers 2.0-era 3.5, Lakers 2.0) total 13.0, the sample's highest, compounded by the benchmark Superteam case (the prime-age 2010 move to Miami), yielding  $\beta_z = +2.99$ , a chasm above second-place Curry (+1.86)—deducted via lambda under the penalty specification and diluted via shares under the Shapley specification (four-title terminal value 4.08  $\rightarrow$  1.75), two philosophies converging. Second, multiple discounts on championship terminal value: two of four titles are asterisk seasons (the 2012 shortened season, the 2020 bubble), all four carry the 2010s Eastern-conference discount, and no title run's opponent average (.579-.674) reaches Olajuwon's 1995 tier. Third, drawdown 0.80, the candidates' highest. Defensively, six All-Defensive selections, but perimeter help and switch ability declined markedly after his peak,  $\alpha_d = -0.17$ .

**Nikola Jokic (-0.56, tied 11th):** offensive alpha in the sample's highest band (playoff BPM 2nd in history, above James and Olajuwon; rTS+ 7.5), with beta as clean as Olajuwon's (the only All-Star on the 2023 champion Nuggets,  $z = -1.86$ ), the exact high-alpha solo-core model. The constraints are sample length (credibility shrinkage 0.86) and championship terminal value (one title, opponent run average .530,  $ATV_z = -1.00$ ). At 29 his inflation-adjusted numbers exceed 29-year-old James across the board, but career depth still needs time.

**Stephen Curry (-2.78, 13th):** the greatest shooter ever, whose gravity rewrote modern spatial concepts; playoff BPM 9th in history, rTS+ 6.5 the sample's second best. His beta is bimodal: the 2017-19 four-All-Star Warriors are a historic Team-beta (load 12.5,  $z = +1.86$ ) requiring heavy deduction; the 2015 and 2022 titles were relatively clean with personal alpha prominent (the 2022 FMVP is the key alpha credential of his no-Durant window). Defensive alpha is the clear weakness (never All-Defensive,  $z = -1.35$ ), with limited individual defensive contribution and downside protection. Four-factor profile: strong offensive alpha, weakest defensive alpha, high beta, ATV marked down by the FMVPs that went elsewhere—about 2.2 standard scores behind the pair above.

Cross-candidate summary: offensive alpha Jokic > James  $\approx$  Curry > Olajuwon; defensive alpha/downside protection Olajuwon >> James > Jokic  $\approx$  Curry; beta cleanliness Jokic  $\approx$  Olajuwon >> Curry  $\approx$  James; sample certainty James confirms with the largest sample, Jokic's is shortest. Gatekeeper = Olajuwon; the tier just behind is Jokic = James (gap

0.03, statistically indistinguishable); then Curry.

## 5.4 Results under Structural Adjustment and the Shapley Specification

**v4 (baseline beta + structural adjustment, theta = 0.4):** Jokic's per-75 real production 26.1 → 23.7 and rTS+ 7.5 → 7.0 (his historic numbers absorb two of the three structural dividends, spacing and assist availability); James 27.7 → 26.0; Curry nearly unmoved (shooter archetype, rim dependence 0.20); Olajuwon and Jordan unchanged (anchor era); Abdul-Jabbar and Johnson gain small physicality-era credit (21.9 → 22.9, 22.8 → 23.3). Gatekeeper race: Olajuwon +0.73 > James -0.78 ≈ Jokic -0.83 > Curry -2.80; the tier holds but the intra-tier order flips to James marginally ahead. **v3 (Shapley shares, lambda = 0):** Olajuwon +0.49 > James -0.18 > Jokic -1.29 > Curry -1.92; under shares, James's and Curry's terminal values are naturally diluted by teammate shares (4.08 → 1.75, 4.05 → 1.98) and Olajuwon (2.47 → 2.06) overtakes—the gatekeeper conclusion holds with no beta penalty coefficient at all. **v4-Shapley (both stacked):** Olajuwon +0.55, all case validations pass. Across the four specifications (v1/v3/v4/v4-Shapley): gatekeeper identical, Duncan > James identical, Jordan first identical.

## 5.5 The Locked Nine, One by One (Factor Decompositions)

**Michael Jordan (+6.46, 1st, by a chasm):** the only player in the framework simultaneously atop all four main channels. Offensive alpha (+1.64) synthesized from per-75 real production 31.7 (sample's highest), playoff BPM 11.1 (first in history), and rTS+ 3.4; defensive alpha (+0.59) backed by 1 DPOY and 9 All-Defensive 1st teams—the guard-position defensive-honor ceiling. All six titles sit in the I = 1.00 nineties, virtually undiscounted ATV with dynasty bonuses on both three-peats, run opponent percentages .582-.686, phi at the 1.25-1.30 tier throughout (six FMVPs); beta mid-pack (load mainly Pippen, 6.53 after league-size normalization); drawdown 0.10 at the sample's floor. His composite leads second place by 3.18 standard scores; first place is invariant under any parameter perturbation (100% in Monte Carlo and LOPO), so using him as one identification condition involves no circularity—the constraint-jackknife test (Section 6.11) shows removing that constraint has zero effect on the feasible region and conclusions.

**Bill Russell (+3.28, 2nd):** the touchstone of the framework's treatment of legacy players. His nominal production (per-75 real production 11.75, the sample's lowest) remains at the tail after Real-Stat calibration in the 1960s' extreme-pace, low-efficiency environment, with negative rTS+ (-2.0)—the model does not look away from his genuine offensive shortfall. Three channels support his seat: the sample's first defensive alpha (consensus-imputed honors 18.0 plus imputed DBPM 5.5, synthesizing +1.62); eleven titles' ATV = 7.65, the sample's second, with the 1959-66 eight-peat earning the

dynasty bonus but every title bearing the deepest small-league discount  $w_{teams} = \sqrt{9/30} \approx 0.55$ —i.e., the model applies the sample's heaviest devaluation to his championships and he still ranks second; drawdown 0.10 and extreme durability. His teammate load (the Celtics' roster depth) is compressed to 3.6 by league-size normalization ( $\times 9/30$ ), justified in Section 4.4: in a 9-team league All-Star slots are relatively abundant and nominal loads overstate true beta.

**Tim Duncan (+2.77, 3rd)**: the largest beneficiary of the risk-adjustment philosophy and the benchmark end of validation case one. Five titles' ATV = 5.28 (the 1999 shortened season honestly discounted at  $w = 0.60$ ; the 2003 title's teammate load only 0.6, with Shapley share 0.77 the highest among multi-title players); beta = 2.50, the locked tier's lowest; defensive alpha (+1.05) from 15 All-Defensive selections plus playoff DBPM 3.9; career drawdown 0.10. What lifts him above Abdul-Jabbar and Bryant is no single peak but the absence of weaknesses: no channel z below  $-0.7$  across all six—the quantitative face of "risk-adjusted superiority."

**Kareem Abdul-Jabbar (+2.06, 4th)**: the exemplar of long-slope stability. Six titles spanning 1971 (a 17-team small league,  $w_{teams} = 0.75$ ) and the Showtime eighties; the latter three titles' teammate loads (Magic + Worthy, 2.8-3.0) push his beta (normalized 7.33) among the locked tier's highest; weighted All-NBA 13.0, the sample's second, earning the omega-channel stability credit second only to James; rTS+ 4.1 and playoff BPM 9.3 (partly an imputed basis for the seventies) support offensive alpha. His profile makes an instructive contrast with James: same ultra-long career and high beta, but Abdul-Jabbar's titles contain no asterisk seasons and no Superteam items, and his defensive depth at defsc = 12.0 is markedly thicker—three differences synthesizing the 2.65-standard-score gap.

**Kobe Bryant (+0.18, 5th)**: factor decomposition in Section 8.2. Supplementary note on v4 (structural adjustment): as a 2000s wing (paint pressure 0.88 era, rim dependence 0.50), his production gains small physicality-era credit and his seat is unchanged. His 0.04 gap to 6th-place O'Neal lies within the noise band, so their relative order is a C-grade conclusion; but the pair's 0.5+ gap to 7th-place Johnson is stable.

**Shaquille O'Neal (+0.14, 6th)**: dominance and shortfall in one profile. Offensive alpha (+0.14) from production 24.95 and rTS+ 3.7; the 2000-02 three-peat at full phi 1.30 (the historic salience of three straight FMVPs) plus the dynasty bonus; beta symmetric with Bryant's (the two ends of one pairing, 8.0 normalized); defensive alpha the locked tier's lowest among bigs ( $-0.49$ ; three All-Defensive 2nd teams plus a rim-presence imputation), drawdown 0.30 carrying the series materialization of the free-throw weakness. The 2006 title as second option (teammate load 2.0, phi = 0.80, East discount 0.93) has its marginal ATV contribution honestly compressed.

**Magic Johnson (−0.36, 7th)**: the first case where the tier rule does work. Offensive

alpha (+0.20) from history's premier assist production (the highest assist share within his 22.84 per-75 real production) plus  $rTS+ 4.5$ ; five titles'  $ATV = 5.14$  with  $\phi$  including the 1.30-tier rookie-year 42-15-7 closeout at center in 1980; but defensive alpha ( $-0.91$ ) is structurally suppressed by the award era (his career overlaps the All-Defensive period but his style never drew defensive narratives), teammate load (Abdul-Jabbar + Worthy) 11.0 the eighties' highest, drawdown 0.45 (the 1984 Finals). His pure score  $-0.36$  sits below the gatekeeper; top-nine eligibility comes from the  $ATV \geq 3.0$  audit rule—flagged D-grade in Section 7.4 and discussed candidly in Section 9.

**Wilt Chamberlain ( $-0.42$ , 8th):** the extreme sample of offensive volume. Under the prime-window basis (1960-68 playoffs, about 28 points and 25.5 rebounds), the 1960s  $I = 1.02$  and per-75 conversion give production 19.76, lifted to 23.17 by structural credit from the paint-pressure 1.16 era; defensive honors consensus-imputed at 12.0 (twice All-Defensive 1st once the award existed). Only two titles (1967 in a 10-team league,  $w = 0.58$ ; 1972 at  $w = 0.75$ ) make his  $ATV = 1.48$  the locked tier's lowest, and drawdown 0.55 carries the historical record of his playoff troughs—the model treats him candidly in both directions: credit for volume and defense, deductions for championship scarcity and volatility, synthesizing  $-0.42$ , with eligibility likewise granted by the audit rule (a Real-Stat dominance chasm).

**Larry Bird ( $-1.59$ , 9th):** the locked tier's lowest score and the most transparent case of the framework's limits. His offensive alpha ( $-0.21$ ) carries upper-middle rather than elite production and efficiency ( $rTS+ 2.4$ , playoff BPM 7.2); defensive alpha ( $-0.40$ ) is suppressed by award thinness (three All-Defensive 2nd teams; DBPM 2.4 partially compensating); three titles'  $ATV = 3.15$  ( $\phi$  including two FMVPs at the 1.25-1.30 tier) just clears the audit line; teammate load (McHale + Parish) 6.90 after normalization; drawdown 0.25. His 2.26-score gap to the gatekeeper is the paper's largest rule-versus-score tension—the third limitation of Section 9 exists precisely for this.

## 5.6 A Unified Reading of the Locked Tier and the Candidate Field

Placing the locked tier and the candidate field side by side, the framework's value geometry becomes visible: the composite top four (Jordan/Russell/Duncan/Abdul-Jabbar) occupy exactly the "championship terminal value  $\times$  low drawdown" double-high quadrant; Bryant and O'Neal represent the "terminal-value driven, efficiency-discounted" quadrant; the gatekeeper Olajuwon represents the "alpha-purity driven, mid terminal value" quadrant—the only player whose candidate-tier terminal value (2.47) pierces the locked tier's score band, his weapons being the sample's cleanest beta and a defensive chasm; James and Curry represent the "high alpha, beta chasm" quadrant, their terminal values diluted by environmental allocation (Shapley) or environmental penalty ( $\lambda$ ); Jokic represents the "alpha extremum, doubly short on sample and terminal value" quadrant. This geometry is topologically invariant across

both beta philosophies (only fine intra-tier ordering adjusts)—direct evidence of the framework's internal consistency.

---

## 6 The Validation System

### 6.0 Design Principles of the Validation System

The thirteen layers are organized by three principles, not piled up. **Principle one (independence)**: validation information is strictly separated from headline conclusions—none of the identification conditions concerns the assignment of seats 10-13; convergent validity uses an out-of-model metric; the divergences in external validity are themselves objects of testing, not fitting targets. **Principle two (adversarialness)**: every layer constructs an opportunity for the framework to fail—ablation deliberately builds crippled models, the constraint jackknife deliberately removes crutches, train/test splitting deliberately creates information barriers, Monte Carlo deliberately injects measurement noise, LOPO deliberately alters sample composition; credibility comes from the survival record under these adversarial settings, not from fair-weather performance. **Principle three (graded reporting)**: results are filed honestly under Section 7.4's invariance grades, passes and failures, strong and weak alike—in particular, the parameter sensitivity of the exact Jokic-James tie (grade C) and the rule-dependent eligibility of three players (grade D) are proactively flagged rather than soft-pedaled.

### 6.1 Case Backtests (Historical Validity)

**Case one (Duncan vs. James)**: +2.77 versus -0.59, a gap of 3.36 standard scores—“decisive edge” reproduced. Attribution: James scores higher on offensive alpha (+0.93 vs. -0.52), but loses at three points—the beta deduction ( $\lambda \times 2.99$ ), the asterisk-season ATV weights (2012/2020), and the drawdown penalty (0.80 vs. 0.10)—while Duncan wins on minimal beta (-0.70), defensive alpha (+1.05), and minimal drawdown, matching the accepted view that “Duncan's career stability and cornerstone role are unmatched” and validating the framework's emphasis on beta deduction and risk control. **Case two (1990s centers)**: Olajuwon +0.67 >> Robinson -2.77 > Ewing -5.06, fully consistent with the historical narrative settled by the 1994-95 head-to-head playoffs (Robinson's complete suppression in the 1995 WCF recorded as a 0.80 drawdown). **Case three (Curry vs. Harden)**: -2.78 versus -6.68, a 3.91 gap; with near-equal offensive indicators, championship terminal value—especially the high-purity-alpha title of the no-Durant window—acts as the tiebreaker, demonstrating ATV's ability to separate statistically similar but achievement-divergent players.

## 6.2 Three-Cutoff Time-Split Forward Tests (Core OOS)

Data are truncated at the ends of the 2016, 2020, and 2022 seasons; the model is rerun with only contemporaneously available information and compared against the full-sample (2025) ranking. "Surprise" = a rank shift exceeding 3 places. Result: **zero surprises at all three cutoffs**. In detail: at the 2020 cutoff Jokic had no title yet, but the model already signaled his rise from historic alpha and minimal beta (offensive-alpha rank 3rd); his subsequent title moved him 13 → 10, a validation of foresight rather than a surprise; James's rank held at 11 across all three cutoffs (by 2020 the model had fully priced his longevity and beta factors; later years grew his total slowly with no jumps—"gliding at altitude"); retired legends shifted zero places, evidence the model rests on stable career-wide evaluation rather than short-term data.

## 6.3 Factor Ablation (Component Necessity)

**Ablating Superteam-beta:** James jumps from 11th to 6th, the gatekeeper gap collapsing from +1.26 to +0.07—model discrimination collapses, in direct conflict with case one's validation, and players succeeding through prime-age stacking are unrealistically elevated. **Ablating inflation deflation** (nominal per-game compared directly): modern candidates systematically rise (Jokic passing the gatekeeper seat, Curry moving up), the Jokic-Olajuwon gap narrowing from 1.23 to 0.78, and the list cannot explain cross-era offensive-defensive environments. Both ablations create severe conflicts with historical validation, proving each component necessary.

## 6.4 Full Parameter Grid Sensitivity (6075 Sets)

Seven parameters (lambda, gamma, mu, omega, rho, st, offense-defense weights) at 3-5 levels each give 6075 sets. The identification conditions of Section 4.8 (margins quantified per the paper's wording: decisive edge  $\geq 1.5$ , decisively ahead  $\geq 1.5$ , significant gap  $\geq 2.5$ ) carve a feasible region of 5402 sets (88.9%). Within it: **the tier conclusion (Jokic and James both above Curry) is 98.2% invariant**; the gatekeeper conclusion holds at 83.2%, decomposing monotonically in lambda—33% at lambda = 0.50, 83% at 0.70, 98% at 0.85, 100% at  $\geq 1.00$ —**with the other six parameters nearly irrelevant at any values**. Across structural intensities theta  $\in \{0.2, 0.4, 0.6\} \times$  both beta philosophies (six combinations), the gatekeeper holds in all, and his margin widens monotonically in theta (+1.39 → +1.51 → +1.63): structural adjustment systematically reinforces rather than shakes the conclusion.

## 6.5 Monte Carlo Uncertainty Quantification (3000 Draws)

All weight parameters perturbed  $\pm 20\%$  uniformly, with Gaussian noise injected into imputation-prone inputs (BPM  $\pm 0.35$ , DBPM  $\pm 0.30$ , defensive score  $\pm 10\%$ , ATV  $\pm 8\%$ , beta load  $\pm 12\%$ , drawdown  $\pm 0.08$ ): P(gatekeeper = Olajuwon) = **97.8%**, P(Duncan >

James) = **100%**,  $P(\text{Jokic and James same tier}) = 60.0\%$ ,  $P(\text{both above Curry}) = **100%**$ ,  $P(\text{Curry} > \text{Harden}) = **100%**$ . Point estimates upgrade to posteriors: James ranking outside the top nine is not the product of a single parameter point but a high-probability event under joint parameter and measurement uncertainty; the "Jokic = James" tie phrasing and the 60% same-tier probability corroborate each other—the gap genuinely lies in the statistically indistinguishable band.

## 6.6 Bayesian Model Averaging

Weighting all 6075 parameter sets by validation likelihood  $w \propto \exp(-2L)$  ( $L$  the hinge loss of the identification conditions): posterior  $P(\text{gatekeeper}) = 82.8\%$ ,  $P(\text{tier}) = 98.3\%$ . BMA replaces "parameter selection" with "probabilistic statements under parameter uncertainty." Because the validation loss is zero over most of the grid, likelihood discrimination is weak and BMA improves little over a uniform prior—the main improvement on lambda sensitivity comes from Section 7.2's structural solution rather than the weighting solution.

## 6.7 Multicollinearity Diagnostics and Factor Dimensionality

Six-factor VIFs:  $\alpha_o = 3.09$ ,  $\alpha_d = 2.69$ ,  $\beta = 2.71$ ,  $\text{ATV} = 2.83$ ,  $\text{AllNBA} = 2.70$ , **media = 6.29**. Media correlates 0.73 with ATV and 0.62 with  $\alpha_o$ —media honors' information is already covered by on-court alpha and championship terminal value, with the lowest independent increment; down-weighting is a statistical necessity. Eigendecomposition of the correlation matrix shows the top three dimensions explain 88% of variance: the six factors effectively span about three independent dimensions (on-court ability / environmental gain / achievement terminal value)—the framework is neither redundant nor incomplete.

## 6.8 Convergent Validity

The model's net alpha is tested against an **out-of-model** independent metric—playoff WS/48 (an entirely different construction, BR basis): Pearson  $r = **0.769**$ . Two independently constructed measures converge strongly; the model's alpha measures the construct "real individual contribution," not modeling noise.

## 6.9 External Validity and Divergence Localization

Spearman rho against media consensus composite ranks (ESPN/The Athletic/HoopsHype) is 0.632; excluding James alone it rises to **0.739**. The model agrees strongly with human consensus, and the sole structural divergence concentrates precisely on the design intent of the beta treatment (James: consensus 2nd vs. model 11th)—a divergence that is not error but an explainable difference in basis: consensus implicitly weights media narrative, which this model explicitly removes. The standard for

an objective model is not full agreement with consensus but the ability to localize and explain every disagreement.

### **6.10 Leave-One-Player-Out (LOPO)**

Each of the 16 calibration-pool players is removed in turn (z-normalization recomputed); the gatekeeper, Duncan > James, and Jordan-first conclusions are checked: **retained in 16/16**. Conclusions do not depend on the presence of any single player (including the composition of the control group).

### **6.11 Constraint Jackknife and Train/Test Constraint Splitting (Added IS/OOS)**

**Constraint jackknife:** removing any one of the five identification constraints and recarving the feasible region: feasible size fluctuates within 88.9%-97.2%, with within-region gate stable at 81.1%-84.7% and tier at 98.2%-98.3%—no single validation case drives the conclusions. **Train/test splitting:** calibrating on cases 1+3 (IS feasible 88.9%), blind-testing on case 2, Robinson > Ewing, and Jordan-first: **OOS pass rate 100%**; the reverse split (calibrate on case 2 + the Ewing ordering, blind-test cases 1, 3, and Jordan) passes at 88.9%. Under both splits, the gatekeeper rate within the OOS-passing sets is 83.2%, identical to the full region—calibration's dependence on either half of the validation cases does not alter the headline conclusions' statistical structure.

### **6.12 Out-of-Sample Player Generalization (6 Players, Frozen Parameters)**

Nowitzki, Garnett, Durant, Antetokounmpo, West, and Malone are scored strictly out-of-sample with normalization frozen on the calibration pool: Antetokounmpo 12/22 (a solo-core champion with two-way alpha, hugging the James tier), West 14, Garnett 17, Nowitzki 18, Malone 19, Durant 20 (the heavy Superteam-beta 2.0 deduction consistent with the framework's axiom). All six land sensibly with zero disturbance to core conclusions—the model, without tailoring to new samples, produces structurally sound output.

### **6.13 Title-Run Opponent Strength: Validating an Independent Information Dimension**

Title-run opponent average win percentage exhibits low correlation with phi (Finals personal performance)— $r = 0.17$  over the full 21-run sample—proving opponent strength is an independent information dimension missing from ATV. Rebuilding ATV-v2 with the multiplier  $w_{opp} = \text{opponent win\%} / 0.60$  (21 title runs recorded from real records, the rest at neutral 0.60): the gatekeeper remains Olajuwon and his lead **widens from 1.23**

to **1.49** (his two runs at .649/.726 are the sample's strongest); Jokic  $-0.67 \approx$  James  $-0.56$  with the tier unchanged; Duncan  $+2.68$  still above James. The new data dimension reinforces rather than shakes the conclusions.

## 6.14 Extended IS/OOS Suite (Seven Added Tests)

**(a) Five-cutoff splits:** adding 2012 and 2024 to Section 6.2's three cutoffs (the 2012 cut drops Jokic and Curry, who had no playoff sample; James per his then-one-title basis), surprises are **0** at both—five historical cutoffs all satisfy the minimal-surprise criterion, with forward validity spanning the framework's entire modern coverage. **(b) Random-subsample cross-validation:** 200 random 12-of-16 sub-pools ( $z$  recomputed per pool); conclusion retention over evaluable pools **151/151 = 100%**—conclusions do not hinge on any particular pool composition. **(c) Placebo/negative controls:** randomly permuting ATV across players 500 times collapses the gatekeeper rate from 97.8% to 45.4%; permuting defensive honors collapses it to 65.4%—conclusions visibly disintegrate when these factors are destroyed, proving they carry real information rather than structural redundancy (had conclusions survived permutation, the model would be data-independent); the collapse is partial because the gatekeeper's beta cleanliness and efficiency advantages persist under permutation, and the collapse magnitudes align with factor weights. **(d) Noise dose-response:** scaling measurement noise from  $0.5\times$  to  $3\times$ , gatekeeper retention runs  $100\% \rightarrow 99.5\% \rightarrow 98.5\% \rightarrow 91.0\% \rightarrow 78.0\%$ —graceful degradation rather than cliff collapse, indicating conclusions rest on multi-factor redundancy, not a single critical point. **(e) Alternative beta measurement (NBA75 basis):** replacing baseline Team-beta loads wholesale with Section 3.5(a)'s NBA75-teammate season counts (league-size normalized): the two bases correlate at  $r = \mathbf{0.884}$  (convergent), and after replacement the gatekeeper remains Olajuwon ( $+0.27 >$  James  $-0.26 \approx$  Jokic  $-0.32 >$  Curry  $-2.30$ ), Duncan  $>$  James, and Jordan-first all hold—beta conclusions are measurement-robust, and on this basis the cleanliness of Olajuwon (2) and Jokic (0) versus the high exposure of James (14) and Curry (12) is reconfirmed by an entirely independent official-list basis. **(f) Weak-helper credit variant:** adding  $+\kappa \cdot z(\text{carry})$  solo-deep-run credit, gatekeeper and case conclusions are unchanged over the entire  $\kappa \in \{0.15, 0.30, 0.45, 0.60\}$  range; the credit lifts Jokic (carry = 2) to  $+0.02$ , markedly closing his distance to the gatekeeper and opening intra-tier distance from James (carry = 1; 2015/2018 excluded under the roster basis), with Olajuwon (carry = 1) himself credited to  $+0.81$ —completing the "weak helper" direction makes the framework's treatment of solo-core players more complete, directionally favoring carry-type players who might previously have been undervalued. **(g) Cross-specification concordance:** Kendall's  $W$  across seven specifications (v1 baseline / v3-Shapley / v4 structural / v4-Shapley / ATV-v2 opponent strength / beta75 basis / carry credit) = **0.971**—all specification variants nearly agree on the 16-player ranking; the framework outputs a stable image of the data structure, not

an accident of specification choice.

## 6.15 Statistical Reinforcement Suite (Five Formal Inferences)

### (a) Specification curve/multiverse + continuous parameter-space check:

continuously and uniformly sampling 3000 v1 and 1500 v3 specifications inside the parameter box (4500 total, simultaneously sealing potential flip pockets between discrete grid points): the gatekeeper margin (Olajuwon – max of other candidates) has sign consistency **90.2%**, median +0.80, IQR [+0.42, +1.09], 5th percentile –0.24 (negatives concentrated in the low-lambda corner, consistent with Section 6.4's lambda curve); the Duncan–James margin has sign consistency **100%**, median +3.45, 5th percentile +2.13. Continuous sampling agrees with the discrete grid, ruling out “conclusions depend on grid-point selection.”

**(b) Permutation hypothesis tests (formal p-values):** under the null that “a factor’s information is exchangeable across players,” 1000 permutations construct the null distribution of the gatekeeper margin: H0[defensive honors exchangeable] **p = 0.048** (rejected), H0[teammate load exchangeable] **p = 0.013** (rejected), H0[championship terminal value exchangeable] **p = 0.321** (not rejected); Fisher’s combined test  $\chi^2 = 17.0$  (df = 6), **joint p = 0.009** (global null rejected). The structure of these p-values itself independently verifies the paper’s factor attribution: the statistically significant load-bearing factors are exactly the two winning channels claimed in Section 5.3—the defensive chasm and beta cleanliness; and the non-significance of the ATV permutation precisely proves the gatekeeper conclusion is **not** a product of ring counting (Olajuwon’s ATV-z is negative; permuting terminal values cannot systematically destroy his margin)—if a ranking framework’s headline conclusion were accused of “counting championships,” this test supplies the formal statistical rebuttal.

**(c) Breakdown analysis (minimal perturbation to flip):** bisection search for flip thresholds on key inputs: Olajuwon’s playoff BPM **cannot flip** the conclusion within its full range ( $\pm 8$ ); raising his drawdown to 1.0 (the theoretical maximum) **cannot flip**; his defensive-honor score must fall by 16.16 (= 2.4 cross-sectional sigma, equivalent to erasing his entire defensive resume) to flip; James’s playoff BPM **cannot flip** within full range upward, and his teammate load must fall by 8.26 (= 2.2 sigma, equivalent to assuming the Heat Big Three and the Lakers duo never existed) to flip; the most sensitive channel is championship terminal value (Olajuwon’s ATV down 1.53 or Jokic’s up 1.51 flips it, each = 0.6 sigma)—a sensitivity direction corresponding exactly to limitation two of Section 9 (the calibration freedom of phi), yet still beyond plausible measurement error in magnitude: 1.53 ATV units roughly equals erasing the entire value of one of Olajuwon’s two titles, or conjuring more than one full-weight championship for Jokic.

**(d) AMIP-style sample breakdown:** greedily removing, one at a time, the non-candidate player most adverse to the gatekeeper margin ( $z$  recomputed each step), it takes  **$k = 4$  removals** (33% of the calibration pool's non-candidate members, the last being Harden) to flip—the conclusion cannot be overturned by small adversarial sample deletion, a progressive step on the same spectrum as Section 6.10's LOPO (16/16 retained at  $k = 1$ ).

**(e) Bootstrap confidence intervals for ranks (rank inference):** 2000 input-noise bootstrap draws; 95% intervals of composite-score rank (16-player pool, controls included): Jordan median rank 1, CI **[1, 1]** (a degenerate interval, no uncertainty at the top); Curry median 13, CI [13, 14]; Olajuwon median 5, CI [5, 7]; James median 10, CI [7, 11]—the upper bound of 7 means at 95% confidence James's composite **never enters the top-six score band**; Jokic median 10, CI [8, 11]; Bryant median 7, CI [5, 10]. Where intervals touch (e.g., Olajuwon's [5,7] and James's [7,11] meeting at 7), pairwise comparison is completed by Section 6.5's pairwise posterior:  $P(\text{Olajuwon} > \text{James}) = 97.8\%$ . Rank intervals upgrade "ranked X-th" from point statements to interval statements—a rare formal-inference form in ranking research.

---

## **7 The Statistical Boundaries of Objectivity: Three-Layer Argument and Explicit Residual Judgments**

"Absolute objectivity" for ranking problems is philosophically unprovable—every scoring function embeds a value function. This paper gives the strongest attainable form of objectivity via a three-layer argument, and discloses its boundaries with equal rigor.

### **7.1 Parameter Independence and the Data Identification of Lambda**

Section 6.4 establishes that **the only sensitive parameter among seven is lambda** (beta-penalty strength); the other six barely affect the headline conclusions at any values. This compresses the framework's entire value judgment into one transparent single-parameter question. Further, lambda is not freely chosen: defining a validation loss  $L(\lambda)$  unrelated to headline conclusions (hinge losses on the identification-condition margins) and marginalizing over the other parameters,  $L(\lambda)$  exhibits a zero-loss plateau on  $\lambda \in [0.7, 1.2]$ , clearly rejecting  $\lambda = 0.3$  ( $L = 0.225$ ) and softly rejecting 0.5 ( $L = 0.003$ )—**lambda's lower bound is identified by the historical validation data**, and the chosen 0.85 lies inside the data-identified optimum. In formal econometric language: the parameter vector is **set-identified** by the moment conditions of the validation cases—the identified set is the feasible region, and headline conclusions are functionals defined on that set; Sections 6.4 and 6.15a show the functional is approximately constant over the identified set (and its continuous

envelope), so the conclusions inherit the full robustness of the identified set without depending on any point within it. The precise proposition is therefore: accepting the framework's beta axiom ("environmental gains warrant substantive deduction," supported by the margin requirements of the historical validation cases), the gatekeeper conclusion holds with 98-100% probability; the tier conclusion (James and Jokic same tier, both above Curry, both below the gatekeeper) holds on 98.2% of the feasible region, nearly independent of any parameter.

## 7.2 Structural Elimination: Shapley Shares Take Lambda to Zero

A stronger treatment makes lambda itself disappear. Specification two (Section 4.5) moves Team-beta from a score-level penalty into ATV-internal credit allocation—under cooperative game theory, a championship's value is allocated to the team's star core by load-proportional approximate Shapley shares, **a rule containing no coefficient**. Rerunning at lambda = 0: gatekeeper = Olajuwon (+0.49), all case validations pass, Jordan first. Residual sensitivity diagnosis (the remaining six parameters, 1215 sets): gatekeeper 87.6%, tier 99.6%, with failures concentrated on a single parameter—the Superteam axiom strength  $st$ , whose flip threshold is  $st \approx 0.4$  (gatekeeper holds at  $st \geq 0.4$ ), the framework taking 0.7 with a 75% safety margin. The per-parameter decomposition gives a finer picture: the gatekeeper rate rises monotonically in  $st$  (70.9% at the 0.5 level, 92.3% at 0.7, 99.5% at 0.9), with secondary sensitivity directions in the career-stability weight  $\omega$  (dropping to 71.4% at the 0.6 level, since it amplifies James's All-NBA depth) and the offensive weight  $w_o$  (75.6% at the 1.15 level)—the precise statement of "failures concentrate on  $st$ " is therefore:  $st$  is the dominant parameter in the flip-threshold sense,  $\omega$  and  $w_o$  constitute second-order sensitivity directions at extreme levels, and only their joint extremes produce  $v_3$ 's failure cases. The framework's value-judgment "surface area" thus shrinks from "a continuous coefficient on all players' total teammate loads" to **one explicit ethical axiom**: actively assembling multi-core structures at one's prime warrants substantive deduction (benchmark cases: James to Miami 2010, Durant to the 73-win Warriors 2016). The extreme test  $st = 0$  (clearing even that axiom) elevates James to gatekeeper—precisely localizing the definitional boundary between the framework and a value system that "draws no distinction in how championships are obtained"; any reviewer can decide for themselves whether to accept the axiom.

## 7.3 Methodological Triangulation of Data-Driven Weights

The entropy-weight and CRITIC methods derive weights entirely endogenously from the data matrix with zero human input. Both output gatekeeper = Olajuwon, with Spearman rho to the calibrated model's ranking of **0.921 and 0.906** respectively, and consistent top-four structure (Jordan / Abdul-Jabbar or Duncan / Russell). Three weight systems of entirely different origin (validation calibration, information entropy, contrast-conflict)

converge to the same ranking structure—measurement-theoretic triangulation: when conclusions are invariant to the weighting method, they reflect the data structure itself rather than the weighter’s intent. The two endogenous weight vectors follow, showing structural closeness to the calibrated weights: the entropy method gives { $\alpha_o$ : 0.141,  $\alpha_d$ : 0.125, anti-beta: 0.089, ATV: 0.198, media: 0.144, All-NBA: 0.152, stability: 0.151}—ATV automatically receives the highest weight, co-directional with the calibrated “championship terminal value as king” ( $\gamma = 2.0$ ); CRITIC gives { $\alpha_o$ : 0.152,  $\alpha_d$ : 0.141, anti-beta: 0.190, ATV: 0.127, media: 0.100, All-NBA: 0.136, stability: 0.154}—anti-beta receives the largest weight for having the highest conflict with other factors, co-directional with the calibrated substantive beta deduction. That is, two zero-input methods each independently “rediscover” the framework’s two core settings.

**The chosen parameter set’s position (the least-commitment point):** normalizing the seven parameters to [0,1] space, the chosen set’s distance to the tightened feasible region’s centroid is 0.062, closer to the centroid than 100% of feasible sets—it is the “least-commitment point” of the model space that passes all historical validation, betting minimally on any single direction. On out-of-sample properties, the chosen set’s time-split surprise count is 0, attaining the feasible region’s global minimum, a minimum jointly attained by 62.8% of feasible sets—the chosen set belongs to the out-of-sample-optimal set rather than being an outlier within it. Honest caveat: the centroid test has a technical weakness—the parameter grid is itself constructed centered on the chosen values, so “near the centroid” is necessary rather than sufficient evidence; the argument’s real force comes from the preceding lambda identification, the invariance grading, and this section’s methodological convergence, with centroid and out-of-sample optimality supplying consistency corroboration rather than independent proof.

## 7.4 Invariance Grading of Conclusions

Synthesizing all validations, conclusions are graded by invariance strength into four levels. **Grade A (parameter invariants, > 98%):** the tier structure—James outside the top nine, same tier as Jokic, both above Curry, Curry significantly above Harden; Duncan decisively above James; Jordan first. **Grade B (axiom-conditional invariants, 98-100% given  $\lambda \geq 0.85$  or the Shapley specification with  $st \geq 0.4$ ):** gatekeeper = Olajuwon, i.e., James ranks 11th rather than 10th. **Grade C (parameter-sensitive, flagged honestly):** the exact Jokic-James tie (median gap 0.96 sigma over the feasible region, Monte Carlo same-tier probability 60%; flipping to James ahead under the Shapley specification)—the cross-specification robust statement is “same tier” rather than “tied.” **Grade D (rule-dependent):** the top-nine eligibility of Johnson, Chamberlain, and Bird is granted by the achievement-audit tier, unsupported by the pure composite score (Section 9). It bears emphasis that the two seats diverging most from media consensus—James 11th and Bryant 5th—are respectively Grade A and a

direct composite-score output: **Bryant's seat depends on no eligibility rule** (his +0.18 ranks 5th within the locked tier by score, synthesized from the dynasty bonus on his five-title ATV, his accumulated defensive honors, and mid-grade alphas across the board), and **James ranking outside the top nine is a Grade-A parameter invariant** (on 98.2% of the 88.9% feasible region his score lies below the gatekeeper's and at least one other candidate's; that he lies below all nine locked players follows automatically at every parameter point where the gatekeeper conclusion holds, since his distance to the gatekeeper exceeds the gatekeeper's distance to the locked tier's boundary).

---

## 8 Discussion and Responses to Objections

### 8.1 The Factor Decomposition of James's Seat: How the Model Derives 11th Place

This seat diverges most from media consensus (2nd) and deserves a full unfolding in the model's language. James's alpha side is impeccable: offensive alpha at the head of the sample's second band, the sample's highest career-stability credit (weighted All-NBA 16.3), four championships of ATV, and peak-years defensive intensity with small-ball-five adaptability. What slides him out of the top ten are three mutually independent, data-backed channels. First, the beta chasm: his contention-window teammate loads (Miami Big Three 4.0/season, Cavaliers 2.0-era 3.5, Lakers 2.0) total 13.0, the sample's highest, compounded by the benchmark Superteam case (the prime-age 2010 move to Miami), so  $\beta_z = +2.99$  stands a chasm above second place Curry (+1.86)—deducted via lambda under the penalty specification, diluted via shares under the Shapley specification (four-title terminal value 4.08  $\rightarrow$  1.75), two philosophies converging on one destination. Second, multiple discounts on championship terminal value: two of four titles are asterisk seasons (2012 shortened, 2020 bubble), all four bear the 2010s Eastern-conference discount, and no run's opponent average win percentage (.579-.674) reaches Olajuwon's 1995 tier. Third, drawdown 0.80, the candidates' highest. After the three channels stack, his composite (-0.59) sits 1.26 standard scores below the gatekeeper (+0.67)—a spacing beyond the Monte Carlo noise resolution threshold, hence  $P(\text{James} < \text{gatekeeper}) = 97.8\%$ . The converse facts must be stated just as plainly: under value systems that strip the Superteam axiom ( $st = 0$ ) or strongly weaken beta ( $\lambda \leq 0.5$ ), James rebounds to the top six or even the gatekeeper seat—the paper does not hide this but presents it explicitly as the definitional consequence of the framework's axioms (Section 7.2).

### 8.2 The Factor Decomposition of Bryant's Seat

Bryant's composite 5th (within the locked tier) is synthesized from four items: five-title ATV = 5.34 (the 2000-02 three-peat earns the  $w\_triple = 1.15$  dynasty bonus; the FMVPs going elsewhere set those three titles'  $\phi$  at 0.85-0.95, the 2009-10 pair at 1.05-1.20), accumulated defensive honors (9 All-Defensive 1st, 3 2nd;  $defsc = 15.75$ , the historical ceiling for guards), career depth at weighted All-NBA 12.8, and low media dependence (MVP+FMVP only 3, a small  $\mu$ -channel share—his standing does not lean on media narrative). His weaknesses are recorded just as faithfully: negative  $rTS+$  (-0.5), the locked tier's lowest playoff BPM (5.9), drawdown 0.45 (the 2004 Finals, the 2010 Game-7 6-for-24), and the Team-beta of the O'Neal partnership. In the model's language Bryant is the archetype "driven by championship terminal value and defensive depth, discounted on efficiency," his +0.18 squeezing in 0.49 below the gatekeeper—if a reviewer asks "Bryant or Olajuwon, who is higher," the honest answer is: their 0.49 gap is inside the Monte Carlo noise band, and the locked-tier eligibility (the ATV chasm of 5.34 versus 2.47) is the decisive ground for Bryant's higher seat—exactly the direct consequence of the framework's "championship as king" axiom ( $\gamma = 2.0$ ).

### 8.3 Anticipated Objections and Replies

**Objection one: "Is the framework tailored to a predetermined conclusion?"** A four-layer reply: (i) calibration uses only four sets of historical validation cases unrelated to headline conclusions, and the train/test blind tests of Section 6.11 (OOS pass rates 100%/88.9%) prove parameters calibrated on either half leak no information about the other; (ii) the headline conclusions' invariance over the feasible region (tier 98.2%) means that even if calibration intent were biased, the parameter space leaves it no room to operate; (iii) two zero-input data-driven weightings independently converge to the same structure (Section 7.3); (iv) the paper proactively reports, under Grade D, the three seats unsupported by score and granted by rule—a conclusion-tailored paper does not voluntarily publish its own failure boundaries. **Objection two: "Defensive honors are also media-voted; why not down-weight them?"** DPOY and All-Defensive teams carry media/coach voting, but unlike MVP their correlation with  $\alpha_o$  is low (defensive contribution escapes box-score-driven narratives);  $\alpha_d = 2.69$  raises no collinearity alarm in the VIF diagnostics, and  $\alpha_d$  blends 50% playoff DBPM, a non-vote metric. **Objection three: "Johnson/Chamberlain/Bird score below the gatekeeper—is the tier rule ad hoc?"** Yes, this is the framework's most substantive rule dependence (Grade D), with identifiable causes: defensive awards only began in 1969 (era-suppressing Johnson's and Bird's  $\alpha_d$ ), and the small-league  $w\_teams$  discount compresses legacy ATV. The paper's treatment is an explicit audit rule plus honest disclosure, not score adjustment to paper over it. **Objection four: "Curry caused the spacing revolution; is adjusting him as a beneficiary a misattribution?"** The endogeneity problem is real and cannot be fully removed; his shooter archetype  $r = 0.20$  already minimizes his exposure (his inputs barely move under the structural

adjustment), and his seat is unchanged before and after. **Objection five: "Jokic's sample is too short."** The 0.86 credibility shrinkage is priced in; the time-split tests show the model signaled his rise before his title; his same-tier relation with James is a Grade-C conclusion, and the paper makes no statement beyond the data. **Objection six: "Doesn't the free-throw intuition contradict your model?"** No—the aggregate data (pace-adjusted, 20% fewer free throws in the 2020s) refute the literal version of "modern soft whistles raise free throws," but support the deeper mechanism (old-era scoring occurred in the foul-rich paint-brawl zone), a mechanism that already enters the model via the paint-pressure index; a separate FT term would double-count.

## 8.4 Falsifiable Forward Predictions

The final test of conditional objectivity is commitment to future evidence. The framework hereby issues four falsifiable predictions, any systematic failure of which should trigger framework revision. **Prediction one:** if Jokic wins another title within three seasons with a solo-core roster (no other first-team-caliber star), the joint growth of his ATV and sample credibility will detach him from the James tier to hold 11th alone; the gatekeeper seat changes hands only if he adds two more titles with at least one run whose opponent strength exceeds .62—the model's forward signal at the 2020 cutoff (offensive-alpha rank 3) is the prior basis of this prediction. **Prediction two:** if Curry adds defensive-team evidence or another high-share championship (teammate load < 1.5) in his remaining career, his 2.2-score distance to the tier above will narrow to within 1.0; otherwise his 13th seat is stable. **Prediction three:** any active player winning a championship via the Superteam path (prime-age assembly, load  $\geq 3.0$ ) will find that title's marginal contribution to historical standing below 45% of an equivalent solo-core championship—a ratio directly computable from the Shapley share formula and verifiable ex post. **Prediction four:** as official tracking data extends backward (archiving shot locations and matchup data), the archetype-imputed rim dependences  $r_i$  will be replaceable with measured values player by player; the framework predicts the shape of the theta sensitivity curve will not change (the gatekeeper's margin still widening monotonically in theta); if measured replacement reverses that monotonicity, the structural-adjustment module should be falsified.

---

## 9 Limitations

First, **imputation dependence:** no BPM before 1974 and no defensive awards before 1969; the relevant values for Russell, Chamberlain, and early Abdul-Jabbar are regression and consensus imputations; Monte Carlo noise covers their magnitude uncertainty but not systematic bias. Second, **the calibration freedom of phi and drawdown:** Finals personal performance and series drawdowns are calibrated on 0.7-

1.3 and 0-1 tiers; though directionally anchored by game records and opponent strength (Section 6.13), the values contain judgment;  $\pm 8\%$  noise sensitivity covers them, and Section 6.15c's breakdown analysis further quantifies the consequence boundary of that freedom: ATV is the only channel flippable within one cross-sectional sigma (0.6), so the calibration of phi and the w-family is the framework's single most adversarially review-worthy link—even though the magnitude required to flip (erasing a full-weight championship) far exceeds reasonable calibration disagreement. Third, **the substantive role of the tier rule**: three players' top-nine eligibility is granted by the audit rule rather than the score (Grade D); the framework's "objectivity" is weaker here than elsewhere, and the paper chooses disclosure over concealment. Fourth, **two axioms resist datafication**: the existence of the beta deduction ( $\lambda > 0$  or the Shapley specification) and the Superteam axiom ( $st \geq 0.4$ ) are normative commitments; data can identify their thresholds and consequences but cannot prove their oughtness. Fifth, **endogeneity**: bidirectional causality runs between Curry and the spacing revolution, and between rule changes and star play styles; treating environment as exogenous in the structural adjustment is a simplification. Sixth, **consensus ranks and the free-throw trend are second-hand compilations**: the external-validity benchmark is a multi-list composite approximation; the free-throw conclusion relies on The Ringer/CTG's basis; the opponent-strength multiplier  $w_{opp}$  is currently recorded from real records for 21 title runs with the rest at the neutral 0.60, and can be formally merged into the main model from the robustness variant once fully recorded. Seventh, **the meta-risk of overfitting to consensus narrative**: the validation cases themselves come from historical consensus; if consensus were systematically biased the framework would inherit that bias—out-of-sample checks (time splits, new players, constraint splitting) mitigate but do not remove this risk. Eighth, **theta and the anchor**: the structural damping and the 1995-96 anchor cannot be uniquely pinned; full-interval sensitivity shows conclusions robust to both, but "robust" is not "independent." Ninth, **single-bucket era assignment for cross-era careers**: each player is assigned one primary era, a simplification for careers spanning structural inflection points (typified by James across the 2012 spacing inflection); Section 4.2 quantifies the simplification's direction (mildly unfavorable to James) and magnitude (a 0.06-sigma blended-bucket correction, conclusions unchanged), and a fully season-weighted era index is the feasible next step in data engineering. Tenth, **the added bidirectional teammate bases are compiled counts**: NBA75-teammate seasons and solo-deep-run counts are cross-computed from public rosters, All-Star histories, and the official NBA75 list; boundary judgments for individual seasons (roster attribution around trade deadlines, counting injured All-Stars) involve basis choices; their use is robustness testing rather than main-score input, and Sections 6.14e/f show conclusions insensitive to them.

## 10 Conclusion

This paper builds and validates a quantitative framework for cross-era NBA player historical standing: inflation deflation and structural era adjustment handle environmental comparability; alpha/beta decomposition and Shapley credit allocation handle the attribution between individual and environment; the six-weight championship terminal value handles achievement quality; drawdown and credibility handle risk and sample; and a thirteen-layer validation system handles credibility. The framework's output ranking: Jordan, Russell, Duncan, Abdul-Jabbar, Bryant, O'Neal, Johnson, Chamberlain, and Bird constitute the top nine; Olajuwon holds seat 10; James and Jokic share the 11-12 tier; Curry sits 13th. Of these, James outside the top nine is a parameter-invariant-grade conclusion (98.2% over the feasible region), Bryant's top-five seat is a direct composite-score output, and the gatekeeper assignment is an axiom-conditional invariant (98-100% under  $\lambda \geq 0.85$  or the Shapley specification). The paper simultaneously provides the conclusion system's complete failure map: under value systems with  $\lambda < 0.7$ ,  $st < 0.4$ , or no distinction in how championships are obtained, James rebounds toward the top; and the tier eligibility of three eighties-and-earlier players depends on the audit rule. We name this form conditional objectivity—axioms made explicit, parameters identified by validation data, sensitivities fully disclosed, conclusions graded by invariance—and argue it is the statistical ceiling of objectivity attainable for historical ranking problems. Future work includes: full recording of title-run opponent strength and the formal merger of  $w_{opp}$  into the main model; player-by-player data validation of post-1997 rim dependence; the re-examination of Grade-C conclusions as Jokic's sample grows; and external-transfer tests of the framework on international basketball samples.

---

## Appendix A Baseline Parameters and Grid

Baseline:  $w_o = w_d = 1.0$ ,  $\lambda = 0.85$ ,  $st = 0.70$ ,  $\gamma = 2.0$ ,  $\mu = 0.25$ ,  $\omega = 0.45$ ,  $\rho = 0.50$ ,  $\theta = 0.4$ ,  $\theta' = 0.5$ , sample credibility (Jokic) = 0.86. Grid:  $\lambda \in \{0.50, 0.70, 0.85, 1.00, 1.20\}$ ,  $\gamma \in \{1.2, 1.6, 2.0, 2.4, 2.8\}$ ,  $\mu \in \{0.15, 0.25, 0.35\}$ ,  $\omega \in \{0.30, 0.45, 0.60\}$ ,  $\rho \in \{0.30, 0.50, 0.70\}$ ,  $st \in \{0.50, 0.70, 0.90\}$ ,  $w_o \in \{0.85, 1.00, 1.15\}$  ( $w_d = 2 - w_o$ ), 6075 sets in total.

## Appendix B Key Data Tables

**B1 Era inflation coefficients:** 1960s  $\approx 1.02$  / 1970s 0.97 / 1980s 1.00 / 1990s 1.00 / 2000-11 0.98 / 2011-20 1.03 / 2020+ 1.07 (anchor 1995-96: ORtg 107.6, Pace 91.8; example 2023-24: ORtg 115.3, Pace 98.5, with 3PA/100 above 17.4 a structural lift). Segment rationale: the 1970s had efficiency slightly below the anchor but faster pace, netting slightly below 1; the 1980s' ORtg sat at  $107 \pm 1$ , near the anchor; the 1990s began

near 108 and fell to about 105 late, overall level with the anchor; 2000-11 was a rule-transition period of high defensive intensity and low possessions (2003-04 as low as about 102.9); 2011-20 saw threes and officiating (freedom-of-movement points) push ORtg to about 110; after 2020 ORtg stays above 112, the highest inflation. Each season's ORtg, Pace, and 3PA/100 are directly recomputable from BR's annual Per 100 Possessions tables. **B2 Structural indices:** paint pressure

1.16/1.16/1.12/1.00/0.88/0.73/0.61; assist availability 0.82/0.88/0.96/1.00/1.02/1.10/1.25

(1960s → 2020s). **B3 Season and league weights:** 1998-99  $w = 0.60$ , 2011-12  $w = 0.80$ , 2020 bubble  $w = 0.86$ ;  $w_{teams} = \sqrt{(teams/30)}$ . Derivation: the 1998-99 season had 50 games, 40% missed →  $w = 0.60$ ; 2011-12 had 66 games, 20% missed →  $w = 0.80$ ; the 2020 bubble regular season totaled 1059 games, about 14% missed, under a neutral-site closed format →  $w = 0.86$ . League-size examples: 1994 and 1995 both had 27 teams,  $w_{teams} \approx \sqrt{(27/30)} \approx 0.95$ ; the 8-14-team Russell era takes the largest discounts (about 0.55 at 9 teams); team counts per era from NBA expansion history and season lists. **B4 Title-run opponent average win percentages (selected):** 1995 Rockets .726, 1997 Bulls .686, 2016 Cavaliers .674, 2011 Mavericks .665, 1993 Bulls .668, 2017 Warriors .622, 1983 76ers .622, 2023 Nuggets .530. **B5 Per-title teammate loads (selected):** 2017/18 Warriors → Curry 4.0; 2012/13 Heat → James 3.2; 2016 Cavaliers → James 2.5; 1994 Rockets → Olajuwon 0.0; 2023 Nuggets → Jokic 0.0; 2003 Spurs → Duncan 0.6; 2011 Mavericks → Nowitzki 0.0.

## Appendix C Code (Unified Single File, Attached)

All computation is integrated into the single file `nba_goat_unified_en.py`, whose internal structure maps one-to-one to the paper's sections: [S3] data (era tables, structural tables, 22 players' data, per-title teammate loads, 21 title-run opponent win percentages, WS/48 and consensus benchmarks) → [S4] methods (inflation index, structural adjustment, alpha/beta, three ATV specifications, two scoring functions, frozen scoring, the parameter grid) → [S5] empirics (`sec5_results`) → [S6] thirteen validation layers (`sec6_cases` / `sec6_timesplits` / `sec6_ablation` / `sec6_grid` incl. jackknife, splitting, BMA, centroid / `sec6_montecarlo` / `sec6_diagnostics` / `sec6_validity` / `sec6_lopo` / `sec6_oos_players` / `sec6_opp` / `sec6_extended` / `sec6_stats`) → [S7] objectivity (`sec7_objectivity`: the lambda loss curve, the Shapley specification, the st threshold, entropy/CRITIC). Running `python3 nba_goat_unified_en.py` prints all results in paper order and generates `results_final_en.csv` and `grid_margins_en.csv`. Randomness is controlled by a fixed seed (7); every number reproduces.

## Appendix D Summary Table of Validation Results

Case backtests 3/3 pass; time splits 2016/2020/2022 surprises 0/0/0; factor ablation

2/2 shows components necessary; grid feasible region 88.9%, tier 98.2%, gatekeeper 83.2% (98-100% at  $\lambda \geq 0.85$ ); Monte Carlo 97.8%/100%/60%/100%/100%; BMA 82.8%/98.3%; VIF media = 6.29; convergent validity  $r = 0.769$ ; external validity  $\rho = 0.632/0.739$ ; LOPO 16/16; constraint jackknife within-region gate stable at 81.1-84.7%; train/test blind tests 100%/88.9%; six out-of-sample players zero disturbance; opponent-strength variant widens the gatekeeper margin 1.23  $\rightarrow$  1.49; the Shapley specification holds at  $\lambda = 0$  with st threshold 0.4; structural adjustment holds over the full theta interval with monotonically widening margin. Extended suite (S6.14): five-cutoff splits surprises 0/0/0/0/0; subsample CV 151/151; placebo permutations of ATV/defensive honors collapse the gatekeeper rate to 45.4%/65.4%; noise dose-response degrades gracefully 100  $\rightarrow$  78%; the beta75 alternative basis at  $r = 0.884$  with conclusions unchanged; carry credit holds over the full kappa interval; Kendall's W across seven specifications = 0.971. Statistical reinforcement (S6.15): multiverse 4500 specifications with sign consistency 90.2%/100% (gatekeeper / Duncan-James); permutation tests defense  $p = 0.048$ , beta  $p = 0.013$ , joint  $p = 0.009$ , ATV permutation non-significant (refuting ring-counting); breakdown—BPM and drawdown unflippable over full ranges, defense needs 2.4 sigma, teammate load 2.2 sigma, the ATV channel most sensitive at 0.6 sigma; AMIP requires adversarial removal of 4 players; rank CIs Jordan [1,1], James [7,11], Curry [13,14].

## Appendix E Notation Table

$I_t$ : era inflation index;  $\mathit{RP}_e$ : *paint-pressure index*;  $\mathit{AE}_e$ : *assist-availability index*;  $r_i$ : *rim dependence*;  $\theta, \theta'$ : *structural damping*;  $\alpha_o, \alpha_d$ : *offensive/defensive alpha*;  $\beta$ : *environmental gain*;  $\mathit{ST}$ : *Superteam flag*;  $\lambda$ : *beta-penalty strength*;  $s$ : *Superteam axiom strength*;  $\gamma$ : *championship terminal-value weight*;  $\mu$ : *down-weighted media credit*;  $\omega$ : *career-stability weight*;  $\rho$ : *drawdown penalty*;  $s_i$ : *sample credibility*;  $w_{\text{season}}, w_{\text{teams}}, w_{\text{triple}}, w_{\text{conf}}, w_{\text{opp}}$ : *season/league-size/dynasty/conference/opponent-strength weights*;  $\varphi$ : *personal real Finals performance*;  $\mathit{ATV}$ : *championship terminal value*;  $\mathit{ATV}^{\text{SH}}$ : *Shapley-share championship terminal value*;  $z(\cdot)$ : *calibration-pool standardization*;  $L(\lambda)$ : *validation hinge loss*; VIF: *variance inflation factor*; LOPO: *leave-one-player-out*; BMA: *Bayesian model averaging*.

## Appendix F Recomputation and Data-Update Guide

Recomputation: `python3 nba_goat_unified_en.py` outputs all of Sections 5-7 in one run (the full table, the gatekeeper race, three specifications, case backtests, multi-cutoff splits, factor ablation, the 6075-set grid with the lambda decomposition, the constraint

jackknife, train/test splitting, BMA, the centroid, Monte Carlo, VIF, convergent/external validity, LOPO, six out-of-sample players, the opponent-strength variant, the L(lambda) curve, the Shapley recheck with the st threshold, and entropy/CRITIC). Environment: Python 3 + numpy/pandas/scipy. Data updates: the file's [S3] block concentrates all inputs—imputed values carry est comments and are individually replaceable; an active player's honor changes need only updates to that player's titles/allnba/media/sample fields; new players are appended in the \_pl() format and assigned to NEW6 (out-of-sample) or ORIG (calibration pool, with justification), then rerun for the full updated conclusion set. Adversarial review suggestion: modify the GRID ranges or the CONS constraint margins and observe whether the invariance conclusions of S6.4 and S7 behave as the paper states.

## **Appendix G Ethics and Interest Statement**

This research involves no human subjects; all data come from public sources; the authors declare the framework's two normative axioms (the beta deduction and the distinction in how championships are obtained) as explicit value commitments, whose alternative-value consequences readers can assess via the sensitivity curves of Section 7; all code is released for adversarial review.

*(End)*