



FORGE Future Work Deep Research

Why the remaining problems point to pillar design and data integrity

Your 2026-02-17 resolution notes make it pretty clear the “weights + calibration” band-aid worked: tier counts came back into range, and Alpha↔PPG rank alignment (Spearman) is strong for RB/TE/WR. The two *remaining* embarrassing inversions, though, are diagnostic gold:

The **RB inversion** (Bucky Irving ¹ over Bijan Robinson ² despite a massive PPG gap) screams that the **Volume pillar is still measuring “how many chances” without pricing those chances correctly**. Fantasy analysts have been yelling “not all touches are created equal” for years, and they’re right: an RB target is worth ~2.5–2.8x a carry in PPR on average, and goal-line / end-zone work is disproportionately valuable.

³

The **QB inversion** (Carson Wentz ⁴ over Josh Allen ⁵) is almost certainly **team context pipeline correctness** (mapping/staleness/join logic), not “model philosophy.” Team context is a legitimate driver of QB scoring because it changes TD opportunity and play volume expectations, which is why Vegas-derived implied points have long been used in fantasy/DFS quarterback selection.

⁶

The future work items you listed are exactly the right next moves. Below is a deep, implementation-minded approach to each one, with the main tradeoffs called out.

Volume pillar redesign that measures quality-weighted opportunity

What “Volume” should mean in a four-pillar system

If you want pillars that stay conceptually clean (and avoid double counting), the sharpest definition is:

Volume = expected scoring from opportunity, independent of player efficiency.

That’s essentially what *expected fantasy points* (xFP) tries to do: strip away talent/efficiency and score only the opportunity profile.

⁷

This framing also aligns with the “weighted opportunity” family of metrics: touches and targets are not equally valuable, and you should weight them based on expected fantasy value.

⁸

Practical spec for a quality-weighted Volume pillar

A production-grade “VolumeScore” can be built from a per-opportunity valuation table plus weekly aggregation.

Step one: build expected points per opportunity type

At minimum for RB/WR/TE (PPR):

- **RB carries**
- baseline carry value
- red-zone carry premium (red zone definition: inside the opponent’s 20) 9
- carry inside the 5 premium (goal-line carries) 10
- **Targets**
- position-specific value of a target (RB targets are worth much more than RB carries in PPR) 11
- add premiums for high-value targets like end-zone targets and deep targets / air yards, which are repeatedly cited as “high value opportunities” in fantasy research 12

This is the core insight: you’re not just counting chances, you’re **pricing** them.

Step two: compute Expected Fantasy Points per game

For each player-week: - $xFP_week = \sum(\text{opportunity_count_type} \times \text{expected_points_value_type})$

Then season level: - $xFP_per_game = \text{mean}(xFP_week)$

Then map to your 0–100 pillar: - $\text{VolumeScore} = \text{normalize}(xFP_per_game \text{ within position and season})$

This makes the Volume pillar far more resistant to the “empty calorie workload” problem (high raw touches, low value touches).

How this fixes the Irving-over-Bijan class of error

A raw-touch system can accidentally treat: - 18 carries between the 20s as similar to - 6 targets + 3 carries inside the 10

...but those are not remotely similar in expected fantasy value. Weighted opportunity/xFP frameworks exist specifically to correct that. 13

So the expected outcome after redesign is that a player whose touches are lower quality (fewer targets, fewer high-value attempts/targets) will stop “hanging” with true elite profiles even if the raw opportunity count looks close.

Keep Volume and Efficiency from cannibalizing each other

Your future work note suggests “blend opportunity count with per-play production rate.” That risks smearing efficiency into volume and then also counting efficiency again in the Efficiency pillar.

A cleaner decomposition is:

- **Volume pillar:** xFP per game (pure opportunity-priced) ⁷
- **Efficiency pillar:** Fantasy Points Over Expected (FPOE) per game, or per opportunity
- ESPN’s xFP framing explicitly pairs xFP with “actual minus expected” style evaluation (they reference opportunity-only and “over replacement” / over expectation concepts in that ecosystem). ¹⁴

This is the “no bullshit” architecture: Volume tells you what the player *should* score on league-average efficiency given their opportunities; Efficiency tells you what they did *above* that.

Stability pillar redesign that measures role consistency, not points smoothness

Why output-variance stability backfires for RB and TE

Fantasy scoring volatility varies substantially by position, with TE repeatedly showing the most weekly variation in at least one large best ball variance analysis. ¹⁵

If “stability” is implemented as “low week-to-week fantasy point variance,” you will systematically reward players who land in the “consistently okay” bucket, and sometimes punish the exact archetype fantasy managers want: high-usage spike-week monsters.

PFF’s own work on consistency makes the key conceptual point: a player can look consistent by scoring the same low number every week, which isn’t actually valuable. ¹⁶

Recommended definition: stability as consistency of opportunity/role

There’s already a well-trodden analytics distinction between **consistency of scoring** and **consistency of usage**. PFF explicitly analyzes “consistency in fantasy usage,” following up their scoring consistency work.

¹⁷

So instead of measuring output variance, redefine Stability by position as role-consistency:

- RB: weekly touch share consistency and snap share consistency
- WR/TE: weekly route participation and target share consistency
- QB: weekly dropback volume consistency (and designed rush share consistency if you track it)

Route participation is well-defined in fantasy analytics as the percentage of team dropbacks where the player ran a route, excluding plays where the player stayed in to block. ¹⁸

A concrete formula that's hard to game

A practical model uses the **coefficient of variation (CV)**, because it scales dispersion by the mean. CV is commonly defined as standard deviation divided by mean. ¹⁹

For a weekly role metric r_w (like route participation or touch share):

- $\text{mean}_r = \text{mean}(r_w)$
- $\text{sd}_r = \text{stdev}(r_w)$
- $\text{cv}_r = \text{sd}_r / \text{mean}_r$ ¹⁹
- $\text{RoleStability} = 1 - \text{clamp}(\text{cv}_r / \text{cv_cap}, 0, 1)$

Then you can blend multiple role metrics: - RB: 60% touch share stability + 40% snap share stability
- WR/TE: 60% route participation stability + 40% target share stability

This gives you "stable opportunity," not "stable fantasy points."

Important guardrail: don't compute stability on tiny roles

CV gets goofy when the mean is close to zero. ²⁰

So you'll want minimum participation gates (per position), such as: - RB: minimum touches per game threshold - WR/TE: minimum routes per game threshold (route participation gives you this naturally) ¹⁸

Without a gate, some random TE2 who runs 6 routes one week and 18 the next will "look volatile" in role, but you shouldn't care because he's irrelevant.

QB team context audit to eliminate Allen-type inversions

Team context is real signal, so data quality has to be unforgiving

Your updated weights moved QB team context up, which is directionally supported by multiple fantasy perspectives:

- Quarterbacks are expected to do better when their team has a high implied point total, because more expected scoring means more TD chances. ²¹
- PFF has also discussed how Vegas-derived team expectations (like projected win totals) can influence fantasy potential and QB scoring environments. ²²

If your pipeline misassigns context (wrong team mapping, stale offense strength, busted join), QB rankings will go completely off the rails—exactly like the Wentz-over-Allen result you called out.

A practical context feature set that's explainable and debuggable

To make context both powerful and diagnosable, bias toward features that have: - clear meaning, - traceable sources, - stable update cadence.

A strong “ContextScore” backbone:

- **Implied team points** (game lines / market expectations) as a scoring environment proxy ²³
- **Pass rate over expected (PROE)** as play-calling tendency proxy (how often a team is choosing to throw relative to expectation) ²⁴
- **Pace / play volume proxy** (even a simplified plays-per-game estimate), because volume creates more attempts and more chances for fantasy points (this is conceptually consistent with the logic behind implied totals and play-calling rates). ⁶

Also: be careful with “QB efficiency” inputs inside team context. A key Sharp Football point is that many efficiency-based metrics are unstable year-to-year, which is a fancy way of saying “this stuff will lie to you if you over-weight it.” ²⁵

Debug checklist for the Buffalo problem

To isolate why the Buffalo Bills ²⁶ context is allegedly wrong, you want hard assertions:

- Team mapping correctness: $\text{player_id} \rightarrow \text{team_id}$ is correct for the season and week (no stale roster join).
- Market data freshness: implied totals / Vegas-derived features are updated for the season week you’re scoring (no off-season cached values). ²⁷
- Pipeline completeness: no NULLs defaulting to league-average or a floor/ceiling that unintentionally crushes elite contexts.
- No double application of “dampening” or “confidence” to the context pillar only (which could pull a great offense down toward baseline more than other pillars).

Given your resolution note explicitly says “team context data issue for BUF,” treat this as a **data incident**: add instrumentation that logs raw inputs for a handful of players/teams into your recompute output so you can see, for example, which implied total and PROE numbers were used. ²⁸

Implementation guardrails that prevent calibration from reintroducing bullshit

Even though your near-term calibration changes widened percentile anchors to reduce amplification, the broader principle is:

Calibration should not invert rank ordering within a position unless you explicitly design it to do so.

If calibration is a monotone mapping (non-decreasing), it cannot flip A>B into B>A (though it can create ties). Isotonic regression is explicitly defined as fitting a non-decreasing function. ²⁹

One caution: isotonic mappings can flatten ranges into ties, and ties can affect rank-based metrics depending on tie-handling. ³⁰

So, regardless of your current approach (percentile remap, anchor-based scaling, etc.), add two guardrails:

- **Monotonicity unit test:** for a large sample of players in the same position/season, assert that if $\text{base_i} > \text{base_j}$ then $\text{alpha_i} \geq \text{alpha_j}$.
- **Amplification test:** assert that a base gap of ϵ cannot become an Alpha gap of $>K$ unless the base is near a boundary and you intend that stretching.

If you ever see “base ordering flipped,” you’ve got either a non-monotone transform or calibration is using additional variables you forgot to account for.

Validation for the future work changes

Your current validation (tier counts, Spearman Alpha↔PPG, anomaly spot checks) is solid. For the future work, add two ranking-centric checks:

- **Top-K recall:** of the top 12 PPG players at each position, what fraction land in T1/T2 after recompute? This is the user-facing “are my elites elite?” test.
- **Inversion penalty:** count inversions where a player with PPG gap $\geq X$ is ranked below a lower-PPG player (separately for top tiers vs whole population). This targets the exact pain you’re trying to eliminate.

Finally, when you redesign Volume to xFP/weighted opportunity, you should expect a structural improvement: weighted-opportunity approaches were created specifically to outperform raw touches as a fantasy predictor because targets and high-value opportunities carry more fantasy expectation. ³¹

1 3 8 11 13 26 31 Weighted opportunity: A better fantasy football predictor ...

https://www.pff.com/news/fantasy-football-weighted-opportunity-a-better-fantasy-football-predictor-than-raw-touches?utm_source=chatgpt.com

2 6 21 23 Using Vegas Lines to Win in Daily Fantasy Football

https://www.si.com/nfl/2017/09/01/fantasy-football-2017-using-vegas-lines-daily-fantasy-football?utm_source=chatgpt.com

4 25 Quarterback Stats That Matter for Fantasy Football

https://www.sharpfootballanalysis.com/fantasy/quarterback-stats-that-matter-fantasy-football-2025/?utm_source=chatgpt.com

5 7 14 Expected fantasy points (xFP): 2025 leaderboard for QBs

https://www.espn.com/fantasy/football/story/_/id/46168860/2025-fantasy-football-expected-fantasy-points-xfp-qb?utm_source=chatgpt.com

9 What are Red Zone Carries? (Fantasy Football)

https://www.fantasypers.com/2023/06/what-are-red-zone-carries-fantasy-football/?utm_source=chatgpt.com

10 12 Fantasy Football: Breaking down high-value touches and ...

https://www.pff.com/news/fantasy-football-breaking-down-high-value-touches-and-targets-for-all-32-nfl-teams?utm_source=chatgpt.com

15 Weekly Variance By Position - A Key To Best Ball

https://underdognetwork.com/football/best-ball-research/weekly-variance-by-position-a-key-to-best-ball?utm_source=chatgpt.com

- 16 Metrics that Matter: Consistency in fantasy scoring, role**
https://www.pff.com/news/fantasy-football-metrics-that-matter-consistency-in-fantasy-scoring-role?utm_source=chatgpt.com
- 17 Metrics that Matter: Consistency in fantasy usage**
https://www.pff.com/news/fantasy-football-metrics-that-matter-usage-consistency?utm_source=chatgpt.com
- 18 What is Route Participation?**
https://www.fantasylife.com/articles/fantasy/what-is-route-participation?utm_source=chatgpt.com
- 19 Coefficient of Variation: Definition and How to Use It**
https://www.investopedia.com/terms/c/coefficientofvariation.asp?utm_source=chatgpt.com
- 20 Coefficient of variation**
https://en.wikipedia.org/wiki/Coefficient_of_variation?utm_source=chatgpt.com
- 22 How Vegas projected win totals can influence fantasy ...**
https://www.pff.com/news/fantasy-football-how-vegas-projected-win-totals-can-influence-fantasy-potential?utm_source=chatgpt.com
- 24 28 Pass Rate Over Expected Data 2025 - Through Week 16**
https://www.draftsharks.com/article/pass-rate-over-expected?utm_source=chatgpt.com
- 27 The Vegas Data You're Not Using, But Should Be**
https://www.draftsharks.com/article/best-worst-offenses-vegas?utm_source=chatgpt.com
- 29 Isotonic regression**
https://en.wikipedia.org/wiki/Isotonic_regression?utm_source=chatgpt.com
- 30 Isotonic calibration changes rank-based test metrics values**
https://github.com/scikit-learn/scikit-learn/issues/16321?utm_source=chatgpt.com