

The Portable Coach:

Which NFL Coaching Effects Transfer Across Rosters, and How Confidently

Nathan Humphrey*

Independent Researcher; Annapolis, MD

ORCID: 0009-0004-8512-4105

May 2026

Abstract

NFL front offices hire and fire coaches as though coaching effects are large, transferable, and known. We show they are none of those things uniformly: coaching authority is *role-specific*, *resolution-specific*, and only partly portable across rosters - and exactly one locus is forecastable precisely enough to price. We apply a single partition (the head-coach / offensive-coordinator / defensive-coordinator staff triple at each team-season) at four resolutions under one signal-to-noise discipline ($B = 1000$ permutation null, cross-season holdout, propensity-matched event design) to nflverse play-by-play 2002–2024. Offense is two-locus: head coaches author macro philosophy (mid-season-firing pass-rate ratio $1.36\times$, $p = 0.002$; off-season pace ratio $1.17\times$, $p = 0.021$, while offensive coordinators author schematic fine-grain off-season red-zone shotgun-rate ratio $1.34\times$, $p = 0.0002$). The decision-relevant quantity is the *predictive-validation slope*: how much of a coordinator’s prior-tenure schematic profile reappears at a new team. Under a hierarchical random-coefficient model the offensive-coordinator slope is 0.469 (95% CI [0.32, 0.62]), robust to team-quality adjustment, and broad - six distinct concepts port at slopes 0.41–0.50. The head-coach slope is 0.279 (95% CI [−0.01, 0.56]), bracketing zero: head-coach portability is real but lives in *which* coach, not a population constant. Under proper role isolation, defensive-coordinator changes move macro tendencies acutely in the modern era (opponent plays $6.37\times$, $p = 0.007$). The same framework applied to the NBA surfaces a different surviving locus (head-coach perimeter defense), demonstrating the method transfers while the sport-specific finding does not. We give the front-office reading throughout and report, rather than hide, the production deployment’s honest detection limits.

1 The front-office decision problem

Coaching is the largest non-player asset a football operation acquires, and the market treats it that way: head coaches and coordinators are hired, fired, and traded on the premise that what a coach

*Contact: mr.nathanhumphrey@gmail.com

did at one team is what he will do at the next. Yet the premise is rarely tested in the terms a front office actually needs. The relevant question is not "do coaches matter" - a season-aggregate fixed-effects question answered affirmatively in prior work - but a pricing question: *which* coaching attributes transfer across rosters, and *how confidently* (Pitts and Evans, 2020, 2019; Bertrand and Schoar, 2003)? A trait that reliably reappears at a new team is an asset a front office can pay for; a trait that is roster-dependent or coach-idiosyncratic is not, however large it looked on the prior team.

This paper reframes coaching effects as a transfer-pricing problem and supplies a calibrated, falsifiable answer. We make five contributions. First, offensive coaching authority is two-locus: head coaches and offensive coordinators occupy distinct, separable resolutions of the same play-calling latent. Second, the offensive-coordinator locus yields a portability slope near 0.47 that is tight, propensity-robust and deployable in a production projection system. Third, the head-coach locus is real but heterogeneous. It should be priced per coach, not as a class. Fourth, under proper role isolation, defensive-coordinator authority is acutely detectable on macro tendencies, overturning a naive "defense is uncoachable acutely" reading. Fifth, the entire apparatus is a framework, not a one-off: applied unchanged to the NBA it surfaces a different surviving locus, isolating what transfers across *sports* (the method) from what does not (the specific finding).

Throughout, we give the decision reading explicitly, and we treat the limits of the evidence (including a production deployment that is below single-season detectability) as results rather than embarrassments.

2 Related work

2.1 Do leaders matter?

The canonical evidence that individual managers imprint a measurable "style" on the organizations they run is Bertrand and Schoar (2003), who identify manager fixed effects across firm policies using a manager-firm matched panel. Sports provide a cleaner laboratory because both the manager's decisions and the organization's output is directly observed (Muehlheusser et al., 2018; Roach and Owens, 2024). Our coach fingerprint clustering (Section 5.4) is the sports instance of the Bertrand-Schoar "style" construct, and our predictive-validation slope is its portability counterpart: not whether a coach has a style, but whether that style travels.

2.2 Coach turnover and the firing event.

A large sports-economics literature finds that managerial dismissal produces, at most, a short-lived bump consistent with regression to the mean rather than a durable improvement (ter Weel, 2011; Bruinshoofd and ter Weel, 2003; Audas et al., 2002; Heuer et al., 2011). This is the correct null for our mid-season-firing event study (Section 5.1): a post-firing shift in team behavior is only informative if it exceeds what mean reversion and roster change would produce on their own, which our continuous-staff permutation control is designed to absorb.

2.3 Closest prior NFL work.

Pitts and Evans (2020) estimate defensive-coordinator and head-coach quality effects on season-aggregate team defense, and Pitts and Evans (2019) estimate offensive-coordinator effects on quarterback performance. These are the nearest antecedents to our Sections 6 and 5.2. We differ in target and unit: where they estimate *quality* fixed effects on season aggregates, we estimate *role-isolated*, *resolution-specific*, and *portability-calibrated* effects (the slope at which a coach’s prior schematic profile reappears at a new team) which is the quantity a front office can actually price.

2.4 Football analytics infrastructure

We build on expected points and play-by-play tooling from the public football-analytics ecosystem, and on the tradition of treating sideline decisions as measurable, sometimes-suboptimal choices (Yurko et al., 2019; Lock and Nettleton, 2014; Carl and Baldwin, 2026; Romer, 2006; Yam and Lopez, 2019).

3 Data

We use nflverse play-by-play for the 2002–2024 seasons joined to Pro Football Reference coaching histories (Yurko et al., 2019; Lock and Nettleton, 2014). The unit of analysis is the staff triple (head coach, offensive coordinator, defensive coordinator) at each team-season. The corpus contains 213 head-coach, 159 offensive-coordinator, and 144 defensive-coordinator off-season changes, plus 45 in-season head-coach firings. All data are free and public; the full pipeline is reproducible from a single orchestration script.

4 Framework and inference discipline

We test one object - the staff triple - at four resolutions: (1) season-aggregate partition tests; (2) within-season event studies around coaching changes; (3) within-game schematic identity by game state; and (4) within-game stress adaptation. Every resolution is held to the same signal-to-noise discipline. We follow a "discipline-pre-registered, architecture-iterative" posture: the inference rules (permutation null, holdout, matching, multiplicity control) were fixed in advance, while which resolution to probe next was decided as each layer’s results motivated the following one. One hypothesis was formally pre-registered before analysis (Section 9).

4.1 Event-study statistic.

For a behavioral cell x (for example, red-zone shotgun rate), we compare the magnitude of change at treated events to change at continuous-staff team-seasons:

$$R = \text{mean}|\Delta x_{\text{treated}}| / \text{mean}|\Delta x_{\text{control}}|$$

with a two-sided $B = 1000$ permutation p -value obtained by reassigning the treated label (Good, 2005). Reporting many cells, we control the false discovery rate per Benjamini and Hochberg (1995); event-study design and inference follow MacKinlay (1997) and the serial-correlation cautions of Bertrand et al. (2004).

4.2 Causal adjustment.

Naive pooled effects are re-estimated under propensity matching on prior team wins, tenure, and year gap, separating roster-confounded shifts from coach-authored ones (Rosenbaum and Rubin, 1983).

4.3 The decision tool: the predictive-validation slope.

For a coach c moving teams, let $\Delta_{\text{pred},c}$ be the schematic shift predicted from the coach’s prior-tenure profile and $\Delta_{\text{obs},c}$ the realized shift. We fit a hierarchical random-coefficient model:

$$\Delta_{\text{obs},c} = \alpha + (\beta + b_c) \cdot \Delta_{\text{pred},c} + \gamma \cdot g_c + \varepsilon_c, \text{ with } b_c \text{ drawn from Normal}(0, \sigma_b^2)$$

where g_c is the year gap and b_c a coach-level slope deviation (Gelman and Hill, 2007). The population slope β is the portability constant; σ_b measures how coach-specific portability is. We estimate the model in PyMC with the No-U-Turn sampler (4 chains, 3000 post-warmup draws, `target_accept = 0.995`)(Abril-Pla et al., 2023; Hoffman and Gelman, 2014). A slope of 1.0 means perfect transfer; 0 means none. This single quantity, with its uncertainty, is what we hand to a front office.

5 Offense is two-locus

5.1 Head coaches author macro philosophy.

At the event-study resolution the head-coach signal is a macro one. In-season firings (the sharpest natural experiment, because the roster is held essentially fixed) shift team pass rate at ratio $1.36\times$ over the continuous-staff control ($n = 45, p = 0.002$). Off-season head-coach changes move pace (plays per game) at $1.17\times$ ($n = 142, p = 0.021$). Both are macro-tendency moves, not fine-grained scheme.

But the decision-relevant question is portability, and here the head-coach signal is more subtle. The population predictive-validation slope is 0.279 with a 95% credible interval of $[-0.008, 0.555]$ - it *brackets zero*. The existence of a head-coach effect survives label-shuffling (Pearson $r = 0.39, p = 0.004$), but it does not survive as a pooled population constant: the credible mass sits on coach-level heterogeneity ($\sigma_b = 0.30$, with posterior probability 1.000 that it exceeds zero). Per-coach posteriors make the point concrete. Some retread hires carry their macro tendency with lower credible bounds above zero - Andy Reid (PHI to KC) at $+0.65 [+0.12, +1.35]$, Frank Reich at $+0.55$, Tom Coughlin at $+0.44$ - while others do not (John Fox $+0.08$, Eric Mangini $+0.10$, Doug

Marrone +0.13). The front-office reading is direct: head-coach philosophy should be priced *per coach*, using that coach's prior record, not as a league-wide "new coach" adjustment. The pooled effect is contaminated by team-mean autocorrelation, exactly the inference hazard Bertrand et al. (2004) warn us against.

5.2 Offensive coordinators author portable scheme.

The offensive-coordinator locus behaves oppositely. It is the one a front office can bank. Off-season coordinator changes shift red-zone shotgun rate at $1.34\times$ ($n = 159, p = 0.0002$). The predictive-validation slope is 0.469 with a 95% credible interval of $[0.320, 0.621]$ - comfortably above zero - with smaller coach-level heterogeneity ($\sigma_b = 0.22$), meaning coordinator portability is more uniform across coordinators than head-coach portability is across head coaches. Crucially, the slope is robust to the team-quality channel that washes out naive pooled effects: adding prior team wins, tenure, and year gap as covariates moves it only from 0.469 to 0.439, and the credible interval continues to exclude zero (Figure 1). This is the difference between a number that survives causal scrutiny and one that does not.

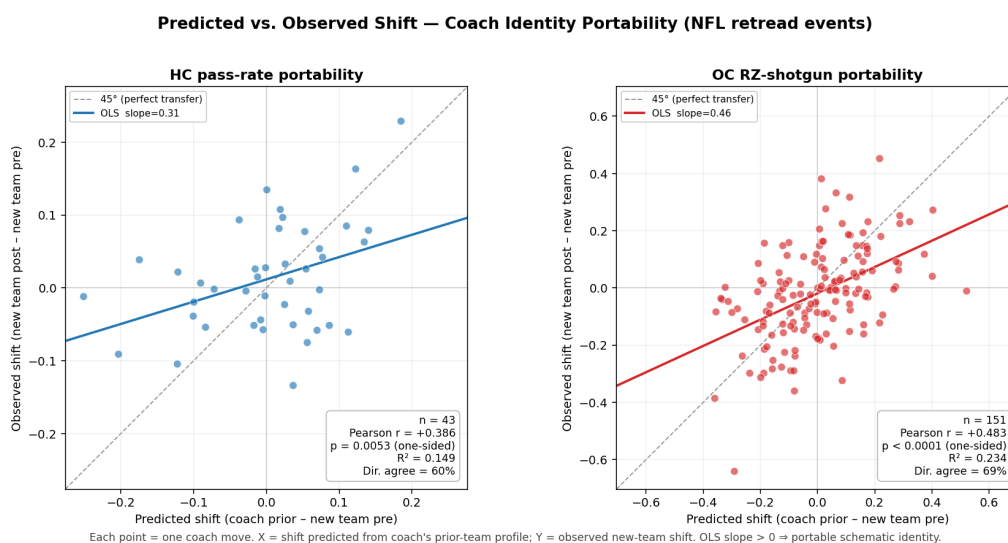


Figure 1: Predictive validation of coaching portability. Each point is a coach who changed teams; the horizontal axis is the schematic shift predicted from the coach's prior-tenure profile, the vertical axis the realized shift at the new team. The offensive-coordinator population slope (0.469, 95% CI $[0.32, 0.62]$) is the deployable portability constant; the head-coach slope (0.279, 95% CI $[-0.01, 0.56]$) brackets zero, indicating coach-level heterogeneity rather than a population mean shift.

5.3 Portability is broad, not narrow.

It would be convenient if a single "signature" concept (red-zone shotgun) carried the coordinator effect. It does not, and the way it fails is informative. We tested six distinct coordinator concepts spanning field-position-conditioned and roster-conditioned families (Figure 2). All six port at OLS

slopes in the 0.41–0.50 band with Pearson r between 0.45 and 0.55. The a priori prediction that field-position-conditioned concepts (red-zone shotgun, goal-line shotgun, two-minute tempo) would transfer more reliably than roster-conditioned ones (deep-shot rate) is *not* supported: deep-shot rate ports at 0.416, statistically indistinguishable from red-zone shotgun at 0.425. The honest reading is that offensive coordinators author a *distributed* schematic identity that travels broadly, not a single portable trick. Two concepts not chosen as the original headline - non-red-zone shotgun rate ($R^2 = 0.30$) and early-down neutral pass rate ($R^2 = 0.30$) - are in fact stronger calibration signals than red-zone shotgun ($R^2 = 0.21$); red-zone shotgun’s appeal is its causal cleanliness (no game-state confounding), not its predictive power. Play-action and run-pass-option rates are unavailable in this play-by-play source and remain untested.

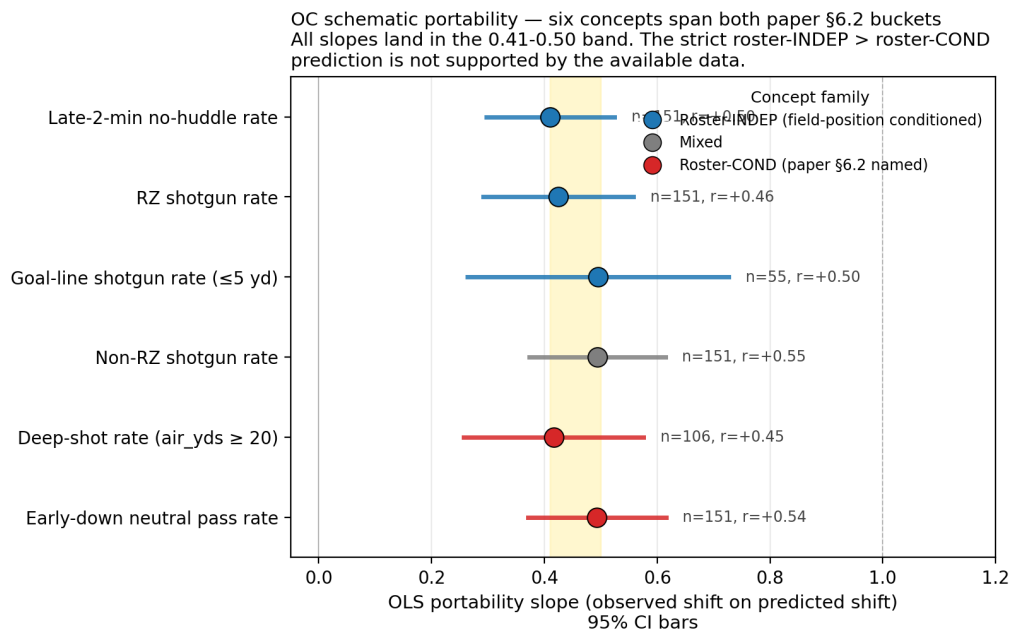


Figure 2: What a coordinator’s whole schematic signature is worth, concept by concept. Each point is the predictive-validation slope (95% CI) at which a coordinator’s prior-tenure rate in that concept reappears at a new team. Because all six transfer at comparable slopes (0.41–0.50), a front office can price a coordinator’s full tendency profile, not just one statistic — the calculation of § 5.2 applies across the board.

5.4 Coaching style is visible in style space.

Clustering coaches in a space of standardized schematic-tendency features recovers interpretable archetypes: four head-coach clusters (silhouettes 0.31-0.34, 129 coaches) and five offensive-coordinator clusters (silhouettes 0.28-0.29, 141 coaches), shown in Figures 3 and 4. This is the Bertrand-Schoar "style" construct made visible for football: coaches supply stable, distinguishable regions of schematic space. Style being visible is necessary but not sufficient for portability - the predictive-validation slope establishes that the style travels.

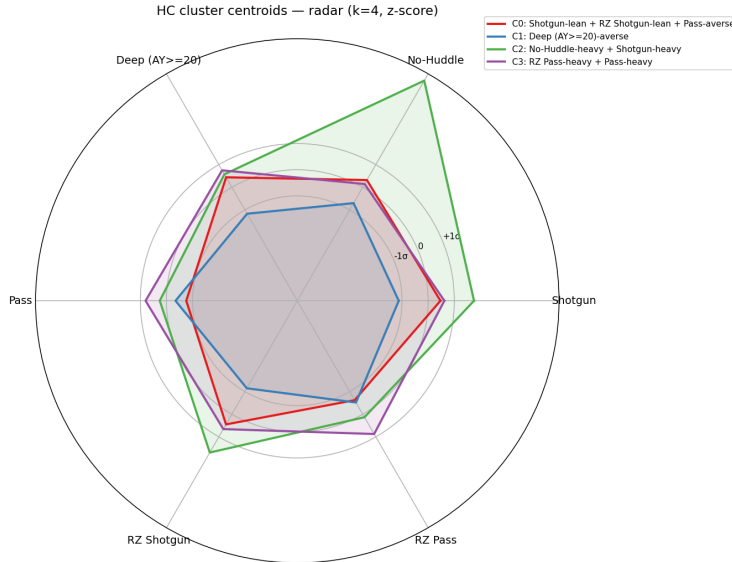


Figure 3: Head-coach schematic archetypes. Each axis is a play-calling tendency (shotgun, no-huddle, deep-throw [air yards ≥ 20], overall pass, red-zone shotgun, red-zone pass), z-scored across coaches and pooled over tenure (games-weighted, ≥ 32 team-games). k -means recovers four silhouette-optimal clusters among 129 head coaches (silhouette 0.34): a pass-heavy/red-zone-pass group (Reid, McCarthy, Payton), a shotgun-lean but pass-averse group (Belichick, Harbaugh, Carroll), a deep-throw-averse traditional group (Fisher, M. Shanahan, Dungy), and a small no-huddle/shotgun-extreme group (Kingsbury, Kelly). Coaches occupy stable, separable regions of schematic space.

6 Defensive authority under role isolation

Defense initially appears to resist acute coaching detection. That appearance is an artifact of mixing event types. When a defensive-coordinator change is pooled with concurrent head-coach changes, the two roles push opposing directions on some cells and the signal cancels. Restricting to *defensive-coordinator-isolated* events (head coach continuous, coordinator changed) recovers acute authority on macro tendencies. In the 2017–2024 cohort ($n = 31$ isolated events), opponent plays shift at $6.37\times$ ($p = 0.007$) and opponent pass rate at $3.94\times$ ($p = 0.050$). Concurrent both-change events ($n = 31$) move opponent plays even harder ($8.99\times$, $p < 0.001$) but are null on opponent pass rate ($1.09\times$, $p = 0.961$) - the cancellation that hid the effect (Figure 5).

The finding is era-textured, and we report the texture rather than the cleanest slice. The 2001–2008 cohort shows the same magnitude on opponent plays ($6.32\times$) but at $p = 0.33$ due to higher within-era variance; the 2009–2016 cohort dips to near-null ($1.25\times$); pooled across 25 seasons the effect is $4.59\times$ at $p = 0.206$. We cannot cleanly judge whether the modern-era sharpness reflects

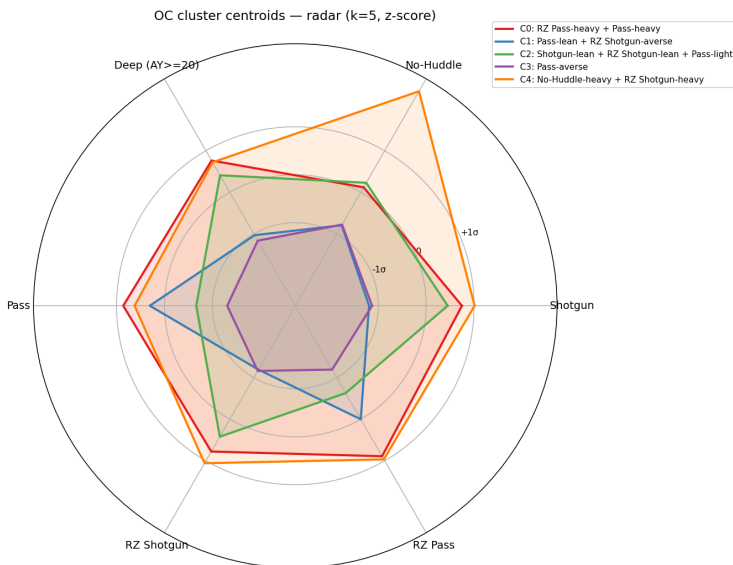


Figure 4: Offensive-coordinator schematic archetypes. Axes and standardization as in Figure 3. Among 141 coordinators, k -means recovers five silhouette-optimal clusters (silhouette 0.29): pass/red-zone-pass-heavy (Bieniemy, Haley, Loggains), pass-lean and red-zone-shotgun-averse (Tom Moore, Arians, Martz), shotgun-lean/pass-light (Bevell, Schottenheimer, McDaniels), uniformly run-heavy "pass-averse" (Kubiak, Cameron, Mularkey), and a no-huddle/red-zone-shotgun-heavy group (Reich, Modkins, O'Brien). The finer five-way split versus the head coaches' four is consistent with coordinators carrying the more granular schematic identity.

a genuine strengthening of the coordinator role (tighter schematic specialization) or simply the best statistical power in our sample; both readings are consistent with the data, and we decline to load-bear on either. At the chronic tenure resolution, coordinator identity is significant on three of seven defensive cells (opponent plays, opponent pass rate, opponent red-zone pass rate; net R^2 between 0.14 and 0.15), rising to five of seven when head-coach identity is added - the same three cells the acute test recovers. Acute and chronic now agree.

7 The framework transfers: a cross-league mirror

The strongest evidence that we have a method, not a coincidence, comes from applying it unchanged to a different sport. Using identical machinery - the same partition, the same permutation null, the same propensity match (with a roster-retention covariate), the same retreat predictive validation - on NBA play-by-play and coaching histories 2014–2024 surfaces a single surviving locus: head-coach perimeter defense (opponent three-point rate). It survives all three rigor layers (naive $1.21\times$, $p =$

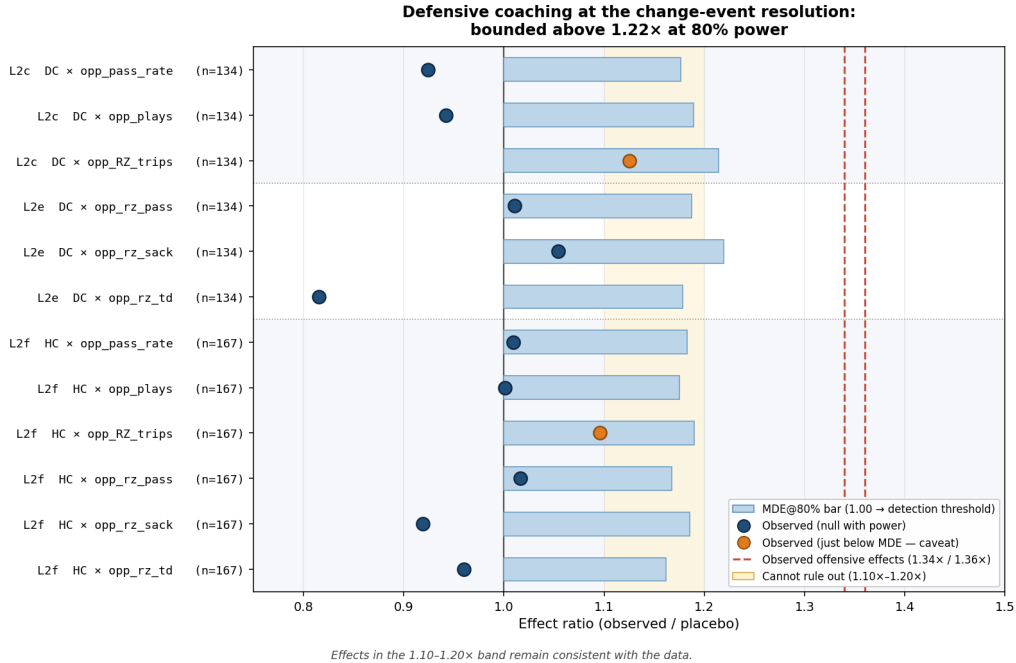


Figure 5: Defensive event-study effects by cell under role isolation. Each row is a defensive macro-tendency; points are magnitude ratios over the continuous-staff control with permutation 95% intervals. Defensive-coordinator-isolated changes (head coach continuous) move opponent plays ($6.37\times$, $p = 0.007$) and opponent pass rate ($3.94\times$, $p = 0.050$) in the 2017–2024 cohort; concurrent head-coach-plus-coordinator changes move opponent plays harder ($8.99\times$) but cancel on opponent pass rate ($1.09\times$, $p = 0.961$). That cancellation is what produced the earlier impression of a defensive null.

0.040; propensity-matched $1.28\times$, $p = 0.048$; reread predictive slope 0.56 , $r = 0.55$, $n = 24$). An NBA *pace* effect that fires naively ($1.20\times$, $p = 0.048$) does *not* survive matching ($1.13\times$, $p = 0.20$). It was roster-confounded, and the same discrimination logic that validates the NFL coordinator slope correctly discards it.

The cross-sport pattern is the contribution. The framework transfers and discriminates scheme from roster in both sports; the *surviving cell* is sport-specific (NFL offensive-coordinator red-zone fine-grain and defensive-coordinator macro tendencies; NBA head-coach perimeter scheme). We offer the sport-geometry difference as an observation, not a mechanism. The NBA analysis is reported in full separately; here it establishes external validity for the method.

8 Application: pricing the coordinator hire

8.1 From slope to decision.

The predictive-validation slope converts a coordinator’s prior-tenure schematic profile into an expected shift at the new team, with calibrated uncertainty. At slope 0.469 , a coordinator who ran red-zone shotgun ten percentage points above league average is expected to move his new team

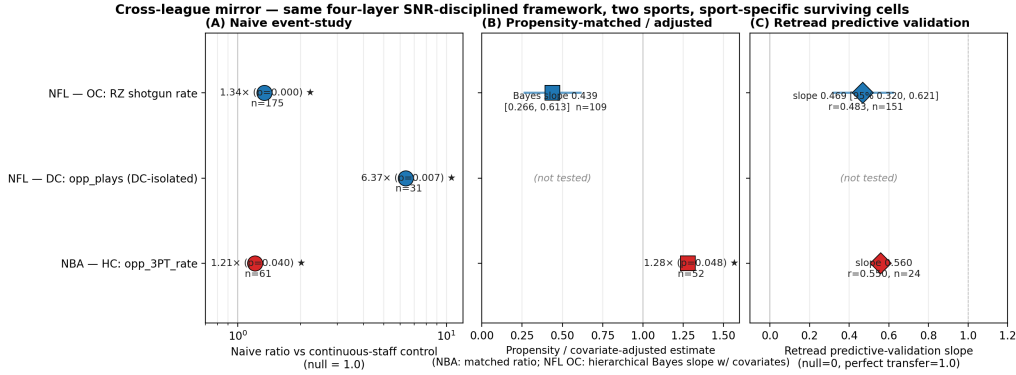


Figure 6: The framework transfers across sports. Each panel applies the identical pipeline—partition, permutation null, propensity match, retread predictive validation—to a different league; the surviving coaching cell differs by sport. In the NFL the offensive-coordinator schematic locus survives (predictive slope 0.47); in the NBA the head-coach perimeter-defense cell (opponent three-point rate) survives all three rigor layers (naive 1.21 \times , $p = 0.040$; propensity-matched 1.28 \times , $p = 0.048$; predictive slope 0.56, $r = 0.55$, $n = 24$), while an NBA pace effect fires naively (1.20 \times) but fails matching (1.13 \times , $p = 0.20$). The method discriminates scheme from roster in both leagues; the surviving cell is sport-specific.

roughly 4.7 points in that direction, with a credible band a front office can attach a confidence to. Because the slope is broad across concepts (Section 5.3), the same calculation applies to a coordinator’s whole schematic signature, not a single statistic. Head-coach philosophy is priced differently: per coach, from that coach’s posterior (Section 5.1, defaulting to no adjustment for first-time hires.

8.2 Production deployment and its honest limits

We wired the offensive-coordinator slope into a production fantasy-projection system (version 1.8) as a pre-week adjustment at off-season coordinator changes. On the 2025 hold-out ($n = 30$ affected players across four teams), the adjustment produced a root-mean-square-error change of -0.005 against the prior version and a direction-agreement rate of 0.533 (bootstrap 95% CI [0.367, 0.700]). This is not a failure of the slope; it is a statement about scale. The implied per-player adjustments (under 1 PPR per season) are two orders of magnitude smaller than typical per-player season projection error (about 130 PPR). A single season therefore cannot empirically confirm or falsify the deployment: the slope’s evidence properly lives at the per-event predictive-validation level. At this level, it is estimated from many retread events of comparable magnitude, not at the per-week production-RMSE level. Direct RMSE confirmation would require roughly ten years of pooled retread events. We deploy the adjustment because it is the empirically calibrated value, and we report its undetectability at single-season scale because hiding it would misrepresent what the data can support.

8.3 What a front office should and should not infer

Three actionable conclusions follow.

1. Offensive-coordinator schematic identity is portable and priceable: a coordinator's prior signature is the best available forecast of his effect at a new team, at a calibrated slope near 0.47.
2. Head-coach philosophy is real but coach-specific: price it from the individual's record, not from a class-level "new head coach" effect. The league-averaged head-coach transfer is statistically indistinguishable from zero.
3. Defensive-coordinator authority on macro tendencies is real and, in the modern era, acutely detectable - but era-textured enough that confidence should scale with recency.

The labor-market implication is that coordinator hires are the more reliably priceable coaching asset, and roster construction should anticipate the schematic shift a portable coordinator brings.

9 Limitations and honest nulls

The one formally pre-registered hypothesis in this program (a per-snap teammate-composition signal check for three skill-position players) returned null at its registered thresholds and is reported as such. The within-game stress-adaptation resolution (Layer 4) was null across all 32 cells. Leave-one-season-out recalibration moves the offensive-coordinator slope by less than 4%. Open confounds remain: quarterback-transition cycles are not separately modeled. Coordinator-inheritance bundles (a head coach who brings in his own coordinator) are only partially disentangled by the joint random-coefficient design; and retread-coach survivorship is unaddressed for lack of an instrument. Play-action and run-pass-option portability are untested for want of data. None of these overturns the two-locus structure or the coordinator slope; all are appropriate next targets.

10 Conclusion

NFL coaching authority is role-, resolution-, and roster-specific. Offense is two-locus: head coaches author macro philosophy that transfers only coach by coach, while offensive coordinators author a distributed schematic identity that transfers broadly at a calibrated, propensity-robust slope near 0.47. Defensive-coordinator authority on macro tendencies is acutely detectable once roles are isolated. The framework that establishes all of this transfers across sports, surfacing a different surviving locus in the NBA under identical rigor. The contribution a front office can use is not a claim that coaches matter, but a calibrated and falsifiable way to price *which* coaching transfers, and how confidently.

10.1 Reproducibility.

Data are public (nflverse play-by-play; Pro Football Reference coaching histories). The full analysis is reproducible from a single orchestration script in the accompanying MIT-licensed package.

References

- Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesebeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., and Zinkov, R. (2023). PyMC: A modern and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, 9:e1516.
- Audas, R., Dobson, S., and Goddard, J. (2002). The impact of managerial change on team performance in professional sports. *Journal of Economics and Business*, 54(6):633–650.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.
- Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics*, 118(4):1169–1208.
- Bruinshoofd, A. and ter Weel, B. (2003). Manager to go? performance dips reconsidered with evidence from dutch football. *European Journal of Operational Research*, 148(2):233–246.
- Carl, S. and Baldwin, B. (2026). *nflfastR: Functions to Efficiently Access NFL Play by Play Data*. R package version 5.2.0.9002.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.
- Good, P. I. (2005). *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Springer, New York, 3rd edition.
- Heuer, A., Müller, C., Rubner, O., Hagemann, N., and Strauss, B. (2011). Usefulness of dismissing and changing the coach in professional soccer. *PLoS ONE*, 6(3):e17664.
- Hoffman, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623.
- Lock, D. and Nettleton, D. (2014). Using random forests to estimate win probability before each play of an NFL game. *Journal of Quantitative Analysis in Sports*, 10(2):197–205.

- MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39.
- Muehlheusser, G., Schneemann, S., Sliwka, D., and Wallmeier, N. (2018). The contribution of managers to organizational success: Evidence from german soccer. *Journal of Sports Economics*, 19(6):786–819.
- Pitts, J. D. and Evans, B. A. (2019). Manager impacts on worker performance in american football: Do offensive coordinators impact quarterback performance in the national football league? *Managerial and Decision Economics*, 40(1):105–118.
- Pitts, J. D. and Evans, B. A. (2020). Defensive coordinator and head coach effects on team defensive performance in the national football league. *Journal of Sports Economics*, 21(5):493–524.
- Roach, M. A. and Owens, M. F. (2024). Updating beliefs based on observed performance: Evidence from NFL head coaches. *Journal of Sports Economics*.
- Romer, D. (2006). Do firms maximize? evidence from professional football. *Journal of Political Economy*, 114(2):340–365.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- ter Weel, B. (2011). Does manager turnover improve firm performance? evidence from dutch soccer, 1986–2004. *De Economist*, 159(3):279–303.
- Yam, D. R. and Lopez, M. J. (2019). What was lost? a causal estimate of fourth down behavior in the national football league. *Journal of Sports Analytics*, 5(3):153–167.
- Yurko, R., Ventura, S., and Horowitz, M. (2019). nflWAR: A reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3):163–183.