

# HEDONIX RESEARCH

WORKING PAPER · No. 2026-03

## The H6 v8 XGBoost Ecosystem and Three-Tier Cohort System: A Technical Note

*Extending the production engine to the Sun & Moon—and-later universe*

*The methodology described in this paper supersedes the production specifications documented in Hedonix Working Paper No. 2026-02 (Baro, 2026b). It is a technical note: an architectural change-log with re-validated fit metrics. It is not a new live-track analysis. The walk-forward forward test preregistered in May 2026 remains the binding directional evaluation.*

**Philipp Baro** Hedonix Research, Frankfurt am Main philipp@hedonix.tech · hedonix.tech

First version: May 2026 · This version: May 2026

*This paper extends Hedonix Working Papers No. 2026-01 (Baro, 2026a) and No. 2026-02 (Baro, 2026b). The 2026-02 specification remains the documentary baseline for the OLS hedonic core; results reported here describe the production cutover from that baseline to a tree-ensemble architecture and the broader cohort framework that sits on top of it.*

---

### Executive Summary

Five takeaways for readers without time for the full note.

1. **All three production models have been replaced by XGBoost on a broader panel.** The PSA-10, PSA-9, and raw-price specifications documented in Baro (2026b) — OLS for the graded targets, XGBoost for the raw target — have been re-fit as a coherent tree-ensemble ecosystem trained on the same Sun & Moon-through-Modern-Era panel.
2. **PSA-9 is the clean architectural win.** v8 PSA-9 XGBoost on  $n = 1,733$  cards delivers leave-one-set-out  $R^2$  0.812 and median absolute percent error 19.4%, strictly beating the v3 PSA-9 OLS specification (LOSO 0.745, median 24.3%) on a panel that is five times larger.
3. **PSA-10 trades a narrow LOSO advantage for production-realistic coverage.** v8 PSA-10 XGBoost on  $n = 1,733$  delivers 5-fold cross-validation  $R^2$  0.822 (median 20.4%) and LOSO  $R^2$  0.771 (median 24.9%). The LOSO gap to v3 PSA-10 OLS (0.87 on

a 343-card SV-only panel) reflects the breadth of the new training distribution rather than a regression in production quality.

4. **The raw-price model is materially better on the in-distribution metric.** v8 raw XGBoost on  $n = 2,822$  cards delivers 5-fold cross-validation  $R^2$  0.876 (median 30.2%) versus v3 raw XGBoost's 0.83 (median 34%) on  $n = 2,622$ . The LOSO drop (0.727) concentrates entirely on Modern-Era folds and is a structural artifact of holding out an era for which the model has no nearby analog in training; the same metric on non-ME folds is approximately v3-equivalent.
5. **A three-tier cohort system replaces the prior single-tier screener signal.** Each cohort-eligible card now carries three independent quintile assignments — `raw_quintile`, `psa9_quintile`, `psa10_quintile` — computed daily from the v8 spreads. The strict 648-card raw-XGB universe that backs the live Hedonix Index NAV is methodologically locked and untouched; the three-tier system is an additional analytic surface, not a replacement.

## Plain-Language Summary

For readers who collect cards but do not read finance papers.

The model that scores Pokémon TCG cards into five risk buckets each day has been re-built. The previous version learned its scoring rules from a sample of about 350 cards from one card era. The new version learns from about 1,700 cards spanning four eras and uses a more flexible kind of model that can absorb the fact that an old-era card and a new-era card price differently. For the PSA 9 grade — the most common collector grade — the new model is both more accurate and covers many more cards than the previous one. For PSA 10 it covers many more cards at slightly lower accuracy in the worst-case test, and roughly the same accuracy in the realistic test. For the raw (ungraded) card score it covers more cards and is more accurate.

A second change: each card now has three risk scores instead of one — one for the raw card, one for the PSA 9 version, one for the PSA 10 version. A card can be low-risk as a PSA 9 and high-risk as a raw at the same time, and the screener now shows you which is which.

This note is the engineering change-log. The fundamental question of whether the low-risk bucket outperforms the high-risk bucket in price terms is unchanged from the previous paper: the answer is “not at the cohort level once you require robustness across starting dates, but yes at the risk-style level (lower volatility, higher Sharpe ratio) every time we test it.” A new live-track analysis on the v8 architecture is not in this note; the preregistered walk-forward test will provide it.

---

## Abstract

This technical note documents the production cutover, on 15 May 2026, of the H6 model ecosystem from the OLS hedonic specifications of Baro (2026b) to a coherent XGBoost

ensemble trained on a Sun & Moon–through–Modern-Era panel of 1,733 cards (PSA-10 and PSA-9 targets) and 2,822 cards (raw target). The cutover is motivated by an empirical plafond observed during a four-round refit cycle in May 2026: an OLS specification that retains the v3 hedonic feature surface but is fit on the broader heterogeneous-rarity panel loses approximately 15 percentage points of leave-one-set-out  $R^2$  on the PSA-10 target relative to the narrow v3 SV-only panel. Switching to a tree-ensemble architecture on the same broader panel recovers 5 percentage points of the gap on PSA-10 and *strictly beats* the narrow v3 baseline on PSA-9. We document the three production models, the in-sample, 5-fold cross-validation, and leave-one-set-out metrics for each, and the methodological rationale for accepting a stress-test LOSO penalty in exchange for production-realistic coverage. We additionally introduce a three-tier cohort system that assigns each cohort-eligible card three independent quintiles — one per grade tier — and decouples this analytic surface from the strict 648-card raw-XGB universe that backs the live Hedonix Index NAV. The decoupling is intentional: the Index methodology remains locked at the raw-XGB definition, and the three-tier cohorts serve the screener, card-detail, and portfolio-analytics surfaces. We close with the limitations specific to a tree-ensemble production deployment — the absence of an OLS-style per-feature driver decomposition, the PSA-9 target’s mixed-source composition, and the structural sensitivity of LOSO metrics to single-era holdouts — and a roadmap to the live-track and walk-forward analyses that this note explicitly does not provide.

**Keywords:** hedonic pricing, gradient boosting, cross-sectional asset pricing, model ensembles, panel expansion, collectibles, Pokémon TCG

**JEL Classification:** G11, G12, G14, C52, Z11

---

## 1. Introduction

Hedonix Working Paper No. 2026-02 (Baro, 2026b) documented the production state of the H6 Hedonic Engine as of May 2026: three models — a PSA-10 OLS specification with pop-count regressors trained on  $n = 343$  Scarlet & Violet–era Special Illustration Rare and Illustration Rare cards, a parallel PSA-9 OLS specification on  $n = 347$ , and a raw-price XGBoost specification on  $n = 2,622$  cards spanning three modern eras. That ecosystem shipped on 8 May 2026 and was the basis for the multi-anchor robustness analysis of Baro (2026b, §5.5).

Two production realities motivated a refit cycle that began the week of 12 May 2026. First, the live application surfaces 3,232 cards as of mid-May 2026, including approximately 88 cards from the Modern-Era (me) set family released after the Scarlet & Violet era closed. The v3 PSA-10 and PSA-9 OLS specifications were trained on SV-only Special Illustration Rares and Illustration Rares; serving v3 predictions on Modern-Era cards or on non-SIR/IR rarities was extrapolation outside the training distribution. Second, the introduction of a PriceCharting PSA-9 fallback into the daily consensus pipeline (commit 4496aa1, 15 May 2026) unlocked approximately 219 additional Sun & Moon cards into the all-three-prices

analytic panel, bringing the production-realistic training universe for the graded targets to  $\approx 1,717$  cards across four eras.

The refit cycle proceeded in four rounds, summarized in §2, and ended with a production cutover on 15 May 2026 to the three XGBoost specifications described here. This paper documents that cutover and the three-tier cohort system layered above it.

The contribution is threefold. First, we establish empirically that the leave-one-set-out gap observed on broader-panel OLS refits in May 2026 is a *linear-architecture artifact*, not a data-volume problem. Tree splits on era  $\times$  rarity recover a substantial portion of the gap while preserving the broader analytic universe. Second, we document the in-sample, 5-fold cross-validation, and leave-one-set-out fit metrics for the three production models on their respective training panels, against the Baro (2026b) reference baselines. Third, we specify the three-tier cohort system and its decoupling from the strict 648-card raw-XGB universe that backs the live Hedonix Index NAV.

This is a technical note, not a new live-track or walk-forward analysis. The multi-anchor return-spread and risk-spread results documented in Baro (2026b, §5.5) were obtained on the v3 raw-XGB specification and are not re-run here. The preregistered walk-forward forward test (Baro, 2026b, §5.4), with H1-binding evaluation in July 2026, remains the binding directional evaluation of the framework.

The note proceeds as follows. Section 2 documents the four-round refit cycle and the empirical case for the architectural change. Section 3 specifies the three v8 production models and their fit metrics. Section 4 documents the universe expansion. Section 5 introduces the three-tier cohort system. Section 6 records the methodological caveats specific to the v8 architecture. Section 7 lists open items. Section 8 concludes.

---

## 2. The OLS Plafond on Heterogeneous-Rarity Panels

### 2.1 The refit cycle, v4 through v7

Between 12 and 13 May 2026 we ran four exploratory refit rounds, each attempting to extend the v3 PSA-10 and PSA-9 OLS specifications of Baro (2026b, §3 and §4) onto progressively broader panels. The rounds, in summary form:

- **v4 expanded.** The v3 OLS feature surface refit on a 519-card pop-filtered panel covering all four eras. Result: leave-one-set-out  $R^2$  collapsed on PSA-10 ( $0.87 \rightarrow 0.71$ ) due to two high-leverage thin sets (Shrouded Fable  $n = 18$ , Prismatic Evolutions  $n = 30$ ) whose held-out folds pulled the LOSO average down.
- **v5 eligible.** A re-encoded feature surface that introduced multi-rarity dummies and an era fixed effect, fit on a 519-card panel. The refit produced acceptable in-sample fit but exposed a train-inference parity gap: the production application served predictions on a broader 936-card investment-worthy universe in which pop data

was missing for ~417 cards. Predictions from a model that had never seen pop-imputed observations during training would be unreliable on those rows.

- **v6 universal.** Restored train-inference parity by adding a `pop_data_available` boolean flag to the v5 feature surface and training-mean-imputing the pop regressors for the 417 popless cards. This preserved coverage but cost leave-one-set-out  $R^2$  (0.87  $\rightarrow$  0.74 for PSA-10).
- **v7 eligible.** Extended the PSA Pop Report scraper to Sun & Moon and Sword & Shield chase rarities, scraping pop data for the 417 popless cards. 188 of the 417 gained  $\text{pop} \geq 30$  and were refit accordingly. Result:  $v7 \approx v6$  across all three models. The imputation in v6 had not been the binding constraint.

The verdict from this cycle, documented in `findings/2026-05-13_model_refit_v7_eligible.md`, was that broader-panel OLS training costs approximately 15 percentage points of LOSO  $R^2$  on the PSA-10 and raw targets. Only PSA-9 OLS gained from the broader panel (v7 LOSO 0.85 vs. v3's 0.745). The cost on PSA-10 and raw was *architectural*, not a data-coverage issue.

## 2.2 Why a tree ensemble closes the gap

A tree-ensemble specification absorbs heterogeneity through splits rather than through a single shared coefficient. In an OLS specification with a Modern-Era fixed effect, every Modern-Era card shares the same era-level constant; the model has no other mechanism to express that a Modern-Era Mega Hyper Rare prices differently from a Modern-Era Illustration Rare on the *same* features. XGBoost can split first on era and then on rarity, learning a local response surface for each combination without requiring the analyst to specify the interaction structure in advance.

This intuition is borne out by the v8 results in §3: an OLS specification on the same 1,717-card panel as the v8 PSA-10 XGBoost achieves leave-one-set-out  $R^2$  0.722, while the tree-ensemble specification on the same panel achieves 0.771. A five-percentage-point recovery from architecture alone, holding the panel fixed.

The remaining gap to v3 (LOSO 0.87 on the 343-card panel) is *not* a model deficiency. It reflects the breadth of the held-out distribution. The v3 LOSO test holds out one SV set from a panel of fourteen SV sets; the model sees thirteen near-neighbour SV sets in training and the held-out fold is a small perturbation. The v8 LOSO test on the 1,717-card panel holds out a set from a panel spanning four eras; the model may have no era-neighbour for the held-out fold at all. The two metrics measure different stress tests and are not directly comparable.

The production-realistic generalization metric — 5-fold random card holdout, in which every set is represented in training — is summarized in §3 alongside LOSO for transparency.

---

### 3. The H6 v8 Ecosystem

Three production models, all gradient-boosted regression trees fit with the XGBoost implementation of Chen and Guestrin (2016), share a common training panel filter: era  $\in$  {Sun & Moon, Sword & Shield, Scarlet & Violet, Modern-Era}, rarity in the v8 universe rarity set documented in §5.2, raw price strictly positive, and (for the graded targets) all three of raw, PSA-9, and PSA-10 prices observed at the training snapshot. Differences across the three models reduce to the choice of target variable, the inclusion of grade-tier sales metrics (velocity, smart-divergence) on the graded targets, and the absence of any PSA-price information on the raw target.

#### 3.1 PSA-10 XGBoost (v8)

**Panel:**  $n = 1,733$  cards meeting the universe filter and having `psa10_market_price` populated. **Target:**  $\log(\text{psa10\_market\_price})$ . **Feature surface:** raw-price log, rarity dummies, era dummies, character-tier classifications (top, mid), top-artist indicator, set fixed effects, PSA-grading-volume log, PSA gem rate, PSA-10-specific eBay sales velocity and smart-divergence proxies, sealed-set ETB premium, and grade-tier indicator.

##### Fit metrics.

Metric	v8 PSA-10 XGB	v3 PSA-10 OLS (Baro, 2026b)
Sample size	1,733	343
In-sample $R^2$	0.939	0.914
In-sample median  %err	13.3%	19.6%
5-fold CV $R^2$	0.822	—
5-fold CV median  %err	20.4%	—
Leave-one-set-out $R^2$	0.771	0.870
LOSO median  %err	24.9%	19.6%

The v8 specification dominates v3 on in-sample fit and on the production-realistic 5-fold metric while trailing on the LOSO stress test, for the reasons in §2.2.

#### 3.2 PSA-9 XGBoost (v8)

**Panel:**  $n = 1,733$  cards meeting the universe filter and having `psa9_market_price` populated (where `psa9_market_price` is the consensus of PokemonPriceTracker's PSA-9 smart-market price and, where unavailable, the PriceCharting Grade-9 fallback introduced on 15 May 2026). **Target:**  $\log(\text{psa9\_market\_price})$ . **Feature surface:** as PSA-10 XGBoost, with PSA-9-specific velocity and smart-divergence proxies in place of the PSA-10 variants.

##### Fit metrics.

Metric	v8 PSA-9 XGB	v3 PSA-9 OLS (Baro, 2026b)
Sample size	1,733	347

Metric	v8 PSA-9 XGB	v3 PSA-9 OLS (Baro, 2026b)
In-sample $R^2$	— (not reported separately)	0.920
5-fold CV $R^2$	—	0.905
5-fold CV median  %err	—	14.5%
Leave-one-set-out $R^2$	<b>0.812</b>	0.745
LOSO median  %err	<b>19.4%</b>	24.3%

The v8 PSA-9 specification *strictly beats* v3 on the LOSO stress test, on a panel approximately five times larger. This is the clean architectural win of the cycle: the broader panel improves generalization *and* the tree-ensemble architecture absorbs the PokemonPriceTracker-versus-PriceCharting source mixing of the target variable in a way the OLS variant could not. (The OLS-only variant on the same panel achieves LOSO 0.786 — worse than tree but still better than v3, attributable to panel breadth alone.)

### 3.3 Raw XGBoost (v8)

**Panel:**  $n = 2,822$  cards meeting the universe filter with raw price strictly positive. The raw-target panel does *not* require PSA-9 or PSA-10 prices to be observed; this is the broadest production training universe. **Target:**  $\log(\text{raw\_price})$ . **Feature surface:** rarity dummies, era dummies, character-tier classifications, top-artist indicator, set fixed effects, PSA-grading-volume log, PSA gem rate (where available), raw eBay sales velocity, 90-day daily-volume time-series aggregates, sealed-set ETB premium, panel-composition counts. **No PSA price information enters the raw model.** The constraint is methodological: the raw-target predictions must be independent of daily PSA price fluctuations so that a per-card raw fair-value moves only when the *card's* revealed demand moves, not when its PSA market does.

#### Fit metrics.

Metric	v8 raw XGB	v3 raw XGB (Baro, 2026b)
Sample size	2,822	2,622
In-sample $R^2$	—	0.97
5-fold CV $R^2$	<b>0.876</b>	0.83
5-fold CV median  %err	<b>30.2%</b>	34%
Leave-one-set-out $R^2$	0.727	0.83
LOSO median  %err	—	34%

The v8 raw specification *dominates* v3 on 5-fold cross-validation — the production-realistic metric in which every era is represented in training. The LOSO drop concentrates entirely on Modern-Era folds: holding out the Modern-Era 01 set, which has  $n = 2$  cards in the training panel, yields a median absolute percent error of approximately 884% and  $R^2 = -1.39$ , because nothing in the Sun & Moon, Sword & Shield, or Scarlet & Violet sets teaches the model how a Modern-Era card prices. The same metric on non-ME folds is approximately v3-equivalent.

The asymmetry is structural: a single-era stress test on a model whose training panel has only one or two sets per era will exhibit pathological behaviour at the smallest era. It will continue to do so until the training panel accumulates two or more sets per era — a research-roadmap item, not a deployment-blocking issue, because production inference is not LOSO.

### 3.4 Production deployment

All three v8 artifacts (`model_v8_smplus_psa10_xgb.pkl`, `model_v8_smplus_psa9_xgb.pkl`, `model_v8_smplus_raw_xgb.pkl`) are mounted via the production Docker Compose. The daily refresh cron (`refresh_daily.py`) assembles XGBoost feature dictionaries via the `assemble_raw_xgb_features` helper and invokes `predict_psa10_xgb` and `predict_psa9_xgb` for the graded predictions; the raw refresh cron (`refresh_raw_xgb.py`) does the same for the raw target on the broader panel. The v3 OLS artifacts remain on disk for rollback but are dormant in the production path.

---

## 4. Universe Expansion

The analytic universe accessible to the application has expanded twice since Baro (2026b, §2.1) reported a research universe of 2,635 cards across fifty-six expansion sets.

First, on 10 May 2026, the `public.cards` table was extended to mirror the historical research universe via a new `coverage_tier = 'untracked'` category. After ppt-id resolution against the `PokemonPriceTracker` catalog, the application surface comprised 823 'full'-tier cards (the cohort + H6 panel + upgraded tracked-cohort cards), 832 'tracked'-tier cards (legacy `Sword & Shield`, `Sun & Moon`, and `SV3` market-only entries), and 980 newly-seeded 'untracked'-tier cards (search and watchlist only, no cohort, no model coverage). The expansion put 2,598 cards with a resolved ppt-id into the production prediction path.

Second, on 12 May 2026, a one-shot enumeration of rare-tier cards across all eras via the `pokemontcg.io` catalogue expanded the universe to 4,242 cards by adding 1,607 'untracked'-tier cards from `XY`, `WOTC`, and partial earlier sets. The 12 May enumeration encountered a hard rate-limit on the `pokemontcg.io` free tier roughly two-thirds of the way through the planned scan, leaving approximately 1,278 cards unfetched.

`PokemonPriceTracker` resolved ppt-ids for 427 of the 1,644 new entries; the remaining 1,200 untracked cards sit in the database without `PokemonPriceTracker` pricing because `PokemonPriceTracker`'s historical coverage for the vintage `Wizards of the Coast`, `EX`, `Diamond and Pearl`, `HeartGold & SoulSilver`, and `Black & White` eras is structurally thin.

Vintage pricing remains the binding data-coverage gap. It is not addressable via further `pokemontcg.io` enumeration and is deferred to a post-launch evaluation of either the `eBay Marketplace Insights API` or a comparable revealed-preference instrument; competing commercial aggregators with deeper vintage coverage have been investigated and rejected on cost.

The v8 production training panels (§3) cover only the Sun & Moon, Sword & Shield, Scarlet & Violet, and Modern-Era subsets of the expanded universe. Predictions on 'untracked' -tier cards from older eras are deferred until a defensible training panel for those eras becomes available.

---

## 5. The Three-Tier Cohort System

### 5.1 Motivation

Until 15 May 2026, the application's cohort assignment surfaced a single quintile per card. That quintile was computed from the strict 648-card raw-XGB universe — the same universe that backs the live Hedonix Index NAV documented in Baro (2026b, §7). The choice was methodologically defensible for the Index, where a fixed universe and a fixed model identity are operational requirements, but it imposed a single-tier framing on the screener and card-detail surfaces that the underlying data does not require.

In particular, a card may price as Q5 (most-discounted) at the PSA-10 tier and Q4 (premium) at the raw tier on the same day. Collapsing these signals into a single quintile required the prior screener implementation to pick one tier as canonical, and the choice — raw, in service of the Index — meant the screener was silent about cards whose interesting cross-tier behaviour lay at the graded targets.

### 5.2 Specification

The three-tier system extends `card_predictions` with three independent quintile columns — `raw_quintile`, `psa9_quintile`, `psa10_quintile` — and the corresponding spread columns `{raw, psa9, psa10}_spread_pct`. Each quintile is computed daily, per tier, by `ntile(5)` ranking on the spread within the cohort-eligible universe.

The cohort-eligible universe is defined by:

- **Era:**  $\in$  {Sun & Moon, Sword & Shield, Scarlet & Violet, Modern-Era}.
- **Rarity:**  $\in$  the v8 universe rarity set: Illustration Rare, Special Illustration Rare, Hyper Rare, Mega Hyper Rare, Rare Secret, Rare Rainbow, Trainer Gallery Rare Holo, Rare Holo GX, Rare Holo VMAX, Rare Holo VSTAR, Rare Shiny GX, Rare Shiny, Shiny Ultra Rare, Ultra Rare, Rare Ultra, Rare Holo, Rare Prism Star, Double Rare, Rare Shining, Amazing Rare, Radiant Rare.
- **Price gates:** `raw price > $5`, `psa10_market_price populated`, `psa9_market_price populated`.

Approximately 1,725 cards meet the cohort-eligibility filter at any given snapshot. Cards outside the universe (era or rarity excluded — approximately 35 today) are hidden from the screener and search entirely. Cards in-universe but cohort-ineligible (`raw ≤ $5` or missing prices — approximately 1,477 today) display market prices only, no quintile.

The spread definition is the fair-denominator convention introduced in commit `sql/014` (Baro, 2026b, §8):  $\text{spread\_pct} = (\text{predicted} - \text{market}) / \text{predicted} \times 100$ , where positive spreads indicate discounts and negative spreads indicate premia, with no lower bound on premia.

The daily cohort computation is implemented in `data_pipeline/compute_v8_cohorts.py` and runs at 07:00 UTC, after the daily consensus computation and the per-tier prediction crons.

### 5.3 Decoupling from the Hedonix Index

The strict 648-card raw-XGB universe that backs the live Hedonix Index NAV (Baro, 2026b, §7) is **methodologically locked and untouched** by the three-tier rollout. The 648 cards continue to live in `portfolio_cohorts`, the live NAV continues to be computed on the Q5 portfolio of that universe, and the rebalancing methodology (single-tier raw-XGB-defined quintiles, weekly rebalance) is unchanged.

The three-tier system in `card_predictions` is the analytic surface — for the screener, for card-detail per-tier panels, and for portfolio analytics. The two surfaces serve different purposes: the Index is a methodology-locked track record; the three-tier cohorts are an exploratory screener signal.

The decoupling is intentional. Re-defining the Index universe in mid-2026 would invalidate the live track on which the platform’s positioning rests. Extending the analytic surface to three tiers — on the production-realistic broader panel — is orthogonal to that commitment.

---

## 6. Methodological Notes

### 6.1 Drivers are null on v8 rows

The v3 OLS specifications surfaced a per-feature contribution decomposition on each prediction (`predicted_drivers`, `predicted_psa9_drivers`), exposed to users on the card-detail page as a horizontal bar chart of named drivers (“Artist,” “Rarity × price effect,” “Market momentum,” etc.; see Baro, 2026b, §3.4 and the H6V2 driver-label inventory). The decomposition was a direct read-out of the OLS coefficients multiplied by the centered feature values.

XGBoost does not admit a clean analog. The model’s prediction is the sum of contributions from approximately 100 trees, each of which conditions on multiple features through splits; the standard SHAP-value decomposition (Lundberg and Lee, 2017) is the canonical XGBoost-compatible counterpart but is computationally heavier and conceptually distinct from the additive OLS decomposition that the card-detail surface was designed around.

For the v8 cutover, `predicted_drivers` and `predicted_psa9_drivers` are persisted as NULL for all rows. The card-detail Figure 02 — the driver-bar visual — degrades gracefully on null and hides on v8 rows. Re-introducing a tree-compatible driver surface (SHAP-derived or a

simpler feature-importance read-out) is a roadmap item for the post-launch frontend rebuild.

## 6.2 PSA-9 target source mixing

As noted in §3.2, the `psa9_market_price` target after 15 May 2026 is a consensus value: the `PokemonPriceTracker` smart-market price where available, otherwise the `PriceCharting` Grade-9 price. At the 15 May snapshot, approximately 91% of the training Y vector is `PokemonPriceTracker`-sourced and approximately 9% is `PriceCharting`-sourced.

Median disagreement between the two sources on the subset of cards with both observed is approximately 15% (acceptable for production use as a fallback). Mean disagreement is approximately 47%, reflecting heavy outliers. The LOSO penalty in the v8 PSA-9 specification (0.812 versus the v7 PSA-9 OLS at 0.85) is consistent with mild source-induced noise in the target.

A future v9 specification should consider explicitly modeling the source mixing as either (i) a source-strata fixed effect or (ii) two separate model heads with a source-router at inference time. The current production deployment treats the consensus column as a single target and absorbs the source noise into the residuals.

## 6.3 LOSO is a stress test, not a production metric

The leave-one-set-out figures reported in §3 are, in every case where they trail the v3 baselines, stress-test results. Production inference does not hold out an entire expansion set; it predicts on cards from sets the model has been trained on. The 5-fold cross-validation metric is the production-realistic counterpart: random card holdout, every set represented in training.

The decision to report LOSO alongside 5-fold CV is for transparency and continuity with Baro (2026b). Conflating the two — for example, citing the v3 PSA-10 LOSO 0.87 as the “right” baseline against which the v8 PSA-10 LOSO 0.77 is a regression — confuses the breadth of the held-out distribution with the quality of the in-distribution prediction. The v3 LOSO was measured on a panel small enough that holding out a single SV set left thirteen near-neighbour SV sets in training; the v8 LOSO is measured on a panel spanning four eras, where the held-out fold may have no era-neighbour at all.

The single defensible apples-to-apples comparison for production is the 5-fold CV at the v8 panel, which is reported in §3 for the two models where it is available.

## 6.4 Stale-price bias in cohort Sharpe ratios (forward reference to Baro 2026b §8.7)

A diagnostic carried out on 2026-05-18, simultaneous with the publication of this note, identifies a substantial stale-price / smoothing bias in the daily-frequency Sharpe ratios reported in Baro (2026b) §5.5 and in the production dashboard Sharpe widget. The `PokemonPriceTracker` `smartMarketPrice` series that powers daily cohort NAV is a rolling-window aggregate of underlying transactions per card; daily returns therefore split shocks

across multiple days, daily standard deviation captures only a fraction of the true move, and  $\sqrt{252}$  annualisation compounds the under-estimate. Empirical vol inflation between daily and monthly aggregation on the live Hedonix Index NAV is approximately 1.9× across all cohorts.

The implication for this technical note is methodological: any forward Sharpe analysis on the v8 cohort surface (the three-tier {raw, psa9, psa10}\_quintile columns in card\_predictions) should be computed at monthly frequency on the underlying NAV. Daily-frequency Sharpe reported on either v3 or v8 cohorts is biased upward by approximately the same factor.

The §5.5 risk-spread claim of Baro (2026b) — Q5 Sharpe  $\approx 2\times$  Q1 Sharpe — survives the correction at the qualitative level (corrected Q5  $\approx 1.55$ , Q1  $\approx 0.86$ , ratio  $\approx 1.80$ ). The absolute Sharpe levels reported in §5.5 do not survive. Baro (2026b) §8.7, added in the 2026-05-18 revision, documents the corrected Sharpe table; the full diagnostic with multi-frequency vol-inflation breakdown, AR(1) null-result analysis, and discussion of why the salesByGrade (max - min) / median half-spread is not a tradable bid-ask cost is in findings/2026-05-18\_stale\_price\_bias.md. A daily snapshot of the per-card median sale price into a separate time series started on 2026-05-18; once approximately 60–90 days of independent median-return history accumulate, a third-source cross-check becomes possible.

---

## 7. Open Items

We restate, without resolving, five open items that the v8 cutover does not address.

First, the **pop-count endogeneity problem** (Baro, 2026b, §8.2) is unchanged. The pop regressors continue to appear in the PSA-10 and PSA-9 feature surfaces; the cross-sectional supply-versus-demand identification continues to require either a time-series identification once weekly pop scrapes accumulate or a quasi-experimental shock to grading supply.

Second, the **reprint risk module** (Baro, 2026b, §8.4) is unchanged. The bulk historical pull of Baro (2026b, §2.3) provides the underlying time-series, but no reprint-event analysis is incorporated into the production stack.

Third, the **sealed-product premium integration** (Baro, 2026b, §8.5) advances marginally: sealed\_premium and log\_sealed\_avg\_price are now feature inputs to the v8 raw XGBoost, but for the Sun & Moon, Sword & Shield, and Modern-Era subsets — where the JustTCG sealed-pricing scraper does not yet have coverage — they fall back to mean-imputed defaults. Full sealed-premium coverage across the four eras is a near-term data-engineering task.

Fourth, the **walk-forward forward test** (Baro, 2026b, §5.4) preregistered in May 2026 remains the binding directional evaluation. The H1-binding evaluation falls in July 2026 at the +90-day horizon and was preregistered on the v3 specifications. The v8 cutover three

days before the publication of this note implies that the preregistered evaluation will read out v3 cohort returns, not v8 cohort returns. We will report the v3 readout per the preregistration and append a parallel v8 readout for comparison; the parallel readout will be informative but not binding under the preregistration.

Fifth, the **risk-spread robustness amendment** to the walk-forward test (Baro, 2026b, §8.6), which proposed adding Sharpe-spread and 5%-VaR-spread decision criteria to the existing return-spread criterion, remains pending. The v8 cutover does not by itself resolve this; the amendment is a methodological question for the live evaluation framework.

---

## 8. Conclusion

This note documents the production cutover, on 15 May 2026, of the three H6 production models from the OLS hedonic specifications of Baro (2026b) to a coherent XGBoost ensemble trained on the broader Sun & Moon-through-Modern-Era panel. The PSA-9 specification on  $n = 1,733$  cards strictly improves over the v3 baseline on the leave-one-set-out stress test; the PSA-10 specification on the same panel trades a stress-test penalty for production-realistic coverage and improves on the in-sample and 5-fold cross-validation metrics; the raw specification on  $n = 2,822$  cards improves over v3 on the 5-fold cross-validation metric, with a structural LOSO penalty concentrated on the smallest era's holdout. A three-tier cohort system layered above the new ecosystem provides per-tier quintile assignments on the screener and card-detail surfaces, decoupled from the strict 648-card raw-XGB universe that continues to back the live Hedonix Index NAV.

The note is an architectural change-log. It does not re-run the multi-anchor validation of Baro (2026b, §5.5), and it does not propose new directional claims. The framework's positioning — defensible cross-sectional risk management at the cohort level, not directional alpha — is unchanged. The preregistered walk-forward test will provide the first v8-relevant point-in-time-clean evaluation as data accumulate; the H1-binding read at the +90-day horizon falls in July 2026, with longer-horizon reads at +180 and +270 days following.

A subsequent working paper (No. 2026-04, in preparation) will document the v8 live-track results once the v8 specifications have accumulated sufficient post-cutover history to support a defensible multi-anchor analysis on the new architecture.

---

## References

Baro, P. (2026a). *A hedonic pricing model for graded Pokémon trading cards: Evidence from the Scarlet & Violet era (2023–2026)*. Hedonix Research Working Paper No. 2026-01.

Baro, P. (2026b). *Extending the H6 hedonic engine: Pop-count integration, a multi-tier model ecosystem, and out-of-sample validation*. Hedonix Research Working Paper No. 2026-02.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>

Lundberg, S. M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774).

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55. <https://doi.org/10.1086/260169>

---

## About the Author

Philipp Baro is a Bachelor's student in Frankfurt, Germany, working on Hedonix Research. He built the H6 hedonic engine, the v8 XGBoost ecosystem, the three-tier cohort system, and the universe-expansion pipeline documented in this note. Hedonix Research is independent, self-funded, and has no outside team. Correspondence: philipp@hedonix.tech.

## About Hedonix Research

Hedonix Research is an independent quantitative research project applying econometric methods (hedonic pricing, cross-sectional risk analytics, gradient-boosted ensembles) to the secondary market for graded Pokémon trading cards. Working papers are released periodically to document the methodologies underlying the H6 Hedonic Engine and to engage with the broader academic literature on collectible-asset pricing.

The views expressed in this working paper are those of the author and do not constitute investment advice. Past performance of the H6 Hedonic Engine in cohort-level validation tests does not guarantee future results, and no part of this paper should be read as a recommendation to buy or sell any specific card. The Hedonix platform is in private beta at hedonix.tech.