

Skeptical review: Statistical Evidence for Coupled Spin-Orbit Evolution in Asteroid Families

Summary

This manuscript tests for population-level signatures of coupled Yarkovsky–YORP evolution in asteroid families by compiling a large merged catalog (~ 1.46 million asteroids from multiple public datasets; Sec. 2.1–2.2), computing family-level aggregate metrics (orbital dispersion, spin-state diversity, characteristic size, age, and counts; Sec. 2.4, Sec. 3.2), and analyzing 50 “well-characterized” families selected by completeness/quality filters (Sec. 2.3–2.4, Sec. 3.1). The main results are (i) a positive association between family age and semimajor-axis dispersion, consistent with cumulative Yarkovsky spreading, and (ii) an additional positive association between semimajor-axis dispersion and spin-period diversity (IQR of spin period), which remains a significant predictor in multivariate regressions controlling for age and median diameter (Sec. 3.3–3.4). The question is timely and the family-aggregate approach is potentially valuable given sparse spin/obliquity data. However, several foundational choices (family definition and whether orbital elements are proper vs. osculating; the use of $\text{Std}(a)$ as a Yarkovsky proxy; and strong selection/sampling biases induced by filtering on scarce spin/obliquity/age data) currently limit the strength of the “coupled evolution” interpretation. Addressing these with clearer provenance, more physically grounded orbital-spreading metrics, explicit uncertainty/sampling-variance propagation, and robustness checks against confounders (belt region, resonances, taxonomy/interlopers) would substantially strengthen both the statistical credibility and the physical interpretability of the claimed coupling signal (Sec. 3.5, Sec. 5).

Strengths

- Clear and well-motivated scientific objective: to search for observable family-level signatures consistent with coupled YORP-driven spin evolution modulating Yarkovsky-driven orbital drift (Sec. 1, Sec. 5).
- Ambitious and useful data compilation effort: programmatic merging of many public datasets into a unified table containing orbital/physical/spin/family/age information (Sec. 2.1–2.2, Sec. 2.6).
- Family-level summary metrics (medians, IQRs, standard deviations) are generally well chosen for robustness to outliers and are coherently reused across the correlation and regression analyses (Sec. 2.4, Sec. 3.2–3.4).
- Use of rank-based correlations and multivariate regression provides an interpretable first-pass statistical test of the central hypothesis (Sec. 2.5, Sec. 3.3–3.4).
- The age–orbital-dispersion relation is recovered as expected, providing an internal consistency check that the pipeline captures a known qualitative trend (Sec. 3.3).
- The manuscript is candid about several limitations (sparse and heterogeneous spin/obliquity/age data; reliance on single-number family descriptors) and frames the work as an initial statistical study rather than a full dynamical model (Sec. 3.5, Sec. 5).

1. **Family definition/membership and the use of proper vs. osculating elements are not specified clearly enough to support family-level dispersion inferences (Sec. 2.1–2.2, Sec. 3.1–3.2). It is unclear (i) what catalog/version and algorithm produced the family assignments, (ii) whether SemimajorAxis_AU/Eccentricity/Inclination are proper or osculating elements, and (iii) whether interlopers/contamination were mitigated. Using osculating elements (or mixed definitions) can artificially inflate within-family dispersions and introduce epoch-dependent noise, directly impacting $\text{Std}_{\text{SemimajorAxis}_{\text{AU}}}$ and related conclusions.**

Recommendation: In Sec. 2.1–2.2 and Sec. 3.1, explicitly document: (a) the family catalog used (name, reference, version/date), family identifiers, and membership definition; (b) whether orbital elements are proper (preferred for family work) or osculating, including the source (e.g., AstDyS/proper-element service) and the exact variables used (proper a , e , $\sin i$ vs i). If osculating elements were used, either redo the primary dispersion metrics with proper elements or provide a quantitative justification/bias estimate (e.g., compare $\text{Std}(a)$ computed from osculating vs proper for a subset of families). Add a brief contamination/interloper discussion: if taxonomic/albedo filtering is not possible, at minimum test robustness by removing obvious outliers in (a, e, i) space or by using published family “core” memberships where available.

2. **The current proxy for Yarkovsky-driven spreading—unconditioned within-family dispersion of semimajor axis (e.g., $\text{Std}_{\text{SemimajorAxis}_{\text{AU}}}$)—is physically coarse and conflates multiple processes (initial ejection field, resonance dynamics, truncation by family-identification boundaries, and size-dependent Yarkovsky V-shapes) (Sec. 2.4.1, Sec. 3.2–3.4, Sec. 3.5). This weakens the causal interpretation of any correlation between orbital dispersion and spin-state diversity.**

Recommendation: Strengthen the orbital-spreading metric in Sec. 2.4.1 and re-evaluate key results in Sec. 3.3–3.4 using at least one more physically grounded alternative: (a) a V-shape slope metric ($|a - a_c|$ vs $1/D$ or H) for each family where sizes/ H are available; (b) size-conditioned dispersion (e.g., $\text{Std}(a)$ computed within narrow diameter bins, then aggregated); and/or (c) $\text{IQR}(a)$ and robust scale estimators with explicit controls for family central a . Demonstrate that $\text{Std}(a)$ tracks (correlates with) the V-shape/Yarkovsky metric for the same families, or explicitly limit the claim to “orbital width” rather than “integrated Yarkovsky drift.” Flag/exclude families strongly affected by nearby major mean-motion/secular resonances (or include resonance-proximity indicators as covariates) and report whether the spin-diversity coefficient remains stable (Sec. 3.4.3, Sec. 3.5).

3. **Filtering on sparse spin/obliquity/age availability likely induces strong selection effects and sampling-variance artifacts (“collider bias”), and the manuscript currently treats family-level spin dispersion (e.g., $\text{IQR}\{\text{SpinPeriod}_{\text{hr}}\}$) as comparably measured across families despite widely varying N and heterogeneous observing strategies (Sec. 2.2–2.3, Sec. 3.1–3.2, Sec. 3.5). With small N (e.g., 5 spin periods), IQR estimates are noisy and can correlate with the number and heterogeneity of observed targets rather than the true family distribution.**

Recommendation: In Sec. 3.1–3.2, report for each of the 50 families: $N_{\text{members,otal}}$, $N_{\text{spin_members}}$, $N_{\text{obliq_members}}$, and (where possible) completeness fractions (e.g., $N_{\text{spin_members}}/N_{\text{members,otal}}$). Add a sensitivity suite in Sec. 3.3–3.4.3: (a) rerun key correlations/regressions at stricter thresholds (e.g., $N_{\text{spin_members}} \geq 20$ and/or $\geq 20\%$ completeness) and compare coefficients; (b) include $N_{\text{spin_members}}$ (and possibly $N_{\text{obliq_members}}$) as explicit covariates; (c) perform within-family bootstrap/resampling of the observed spin sample to estimate uncertainty on $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ and propagate it into an errors-in-variables or weighted regression. Summarize the domain of validity in Sec. 3.5/5 (e.g., “large, well-observed families with adequate spin sampling”).

4. **Family_Age_Gyr is central to the analysis but its provenance, construction, and uncertainties are under-specified and not propagated (Sec. 2.4.4, Sec. 3.2–3.4). The text implies member-level ages aggregated to a family-level age when “varied significantly,” which is conceptually unclear for family ages and risks introducing non-reproducible branching logic and biased regression coefficients.**

Recommendation: In Sec. 2.4.4, provide a reproducible age model: list the age source(s) with references/versioning; clarify whether ages are truly per-family (preferred) or per-object; and define an explicit, deterministic aggregation rule if multiple literature ages exist per family (e.g., median of published estimates, with a recorded spread). Add an uncertainty field per family (quoted literature uncertainty or inter-estimate spread) and propagate it in Sec. 3.3–3.4 via (a) Monte Carlo perturbation of ages within uncertainties, and/or (b) down-weighting/removing families with large age uncertainty. Report how often the key $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ coefficient remains positive/significant under these perturbations (Sec. 3.5).

5. **Statistical robustness is not yet demonstrated at a level commensurate with the small sample size (50 families) and the number of explored correlations/models (Sec. 2.5, Sec. 3.3–3.4). Key gaps include: incomplete multiple-testing control for the correlation matrix; limited influence diagnostics; and limited reporting of effect-size uncertainty (confidence intervals) for the main regression coefficient interpreted as the coupling signature.**

Recommendation: In Sec. 2.5 and Sec. 3.3–3.4, explicitly enumerate (a) the set of confirmatory hypotheses (pre-specified) versus exploratory tests; (b) the number of correlation pairs examined; and (c) the regression models compared. Apply a concrete multiple-testing procedure (Holm or FDR) to the correlation results used for claims (Table 1 and any supporting text) and report adjusted p -values. For the main regression (Eq. (2) and any alternatives), add: bootstrap or robust-SE confidence intervals; leverage/Cook’s distance and a leave-one-family-out (LOO) plot of the $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ coefficient; and a robust fit (Huber RLM / Theil–Sen) to show sign/magnitude stability. Include a compact diagnostics figure/table (residuals, heteroskedasticity test, influential points) for the primary model (Sec. 3.4.3).

6. **The interpretation of a $\text{Std}\{\text{SemimajorAxis}_{AU}\}$ – $\text{IQR}\}$ association as evidence for YORP \rightarrow Yarkovsky coupling is vulnerable to confounding by unmodeled family/environment properties: belt region (heliocentric distance), thermal/compositional class, resonance proximity, family identification quality/contamination, and observational targeting (Sec. 3.4.3, Sec. 3.5, Sec. 5). These factors can jointly affect both measured orbital width and measured spin diversity without requiring dynamical coupling.**

Recommendation: Temper causal language throughout (Abstract, Sec. 1, Sec. 3.4.3, Sec. 5) to emphasize “consistent with” rather than “evidence for” coupling unless robustness checks support stronger phrasing. Then, in Sec. 3.4.3, add controls/proxies for key confounders: family central semimajor axis (or belt region category: inner/mid/outer), inclination/eccentricity regime, resonance-proximity flags, and (where available) taxonomic/albedo indicators. Refit the main model with these added covariates and report whether the spin-dispersion term remains stable. If taxonomy is unavailable broadly, include at least NEOWISE albedo where present or restrict to families with consistent albedo/type as a robustness subset.

7. **Spin-state characterization is concentrated on a single metric ($\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$), while obliquity—more directly tied to the sign/magnitude of Yarkovsky drift—is underdeveloped and likely too sparse for stable inference at current thresholds (Sec. 2.4.2, Sec. 3.2, Sec. 3.4.2). In addition, spin-period dispersion in linear hours is sensitive to long-period tails; a log-period dispersion may be more appropriate and comparable across families.**

Recommendation: In Sec. 3.3–3.4.2, present a systematic metric comparison: repeat the key regression(s) using (a) $\text{Std}\{\text{Log_SpinPeriod}\}$ (with a precised definition; see minor issues), (b) $\text{IQR}\{\log P\}$, and (c) any feasible obliquity metrics. For obliquity, either (i) explicitly limit/withdraw obliquity-based conclusions due to small n and treat results as exploratory, or (ii) restrict to a higher N threshold and use metrics better matched to Yarkovsky expectations (e.g., fraction near 0° – 180° , bimodality indicators, or prograde/retrograde split if sense can be inferred). Report null results clearly as constraints rather than omitting them.}

8. **The paper reports fitted slopes (e.g., age–dispersion and spin-IQR coefficients) but does not quantitatively connect them to plausible Yarkovsky/YORP magnitudes across the sampled size and heliocentric-distance range, nor does it benchmark against alternative mechanisms (Sec. 3.4.1–3.5, Sec. 5). This makes it difficult to assess whether the estimated effect sizes are physically reasonable.**

Recommendation: In Sec. 3.5, add an order-of-magnitude physical check: compute representative Yarkovsky drift rates (as a function of D , a , and obliquity assumptions) and compare integrated drift over 0.1–3 Gyr to the observed family-level dispersion scale. Then interpret the regression coefficient for $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ (or log-spin metric) in terms of plausible YORP-driven spin/obliquity diversification and its impact on the distribution of drift rates. Explicitly discuss alternative drivers (collisional reorientation, resonance-driven spreading, family truncation/contamination) and include at least one simple exclusion/control test (e.g., remove resonance-adjacent families) to show these are unlikely to fully explain the pattern.

9. **Data provenance and catalog-level systematics are not described in sufficient detail to evaluate biases and ensure reproducibility (Sec. 2.1–2.2, Sec. 2.4.4, Sec. 2.6). The manuscript references internal filenames (e.g., `asteroid_age.csv`) without clear mapping to published sources/versions, and it is unclear which spin/obliquity catalogs (and methods) dominate the final sample.**

Recommendation: Add a concise provenance table in Sec. 2.1–2.2 and/or Sec. 2.6 listing each key quantity (family membership, proper elements, diameters/albedos, spin periods, obliquities, ages) with: source catalog/survey, literature reference, version/access date, approximate coverage, and known selection effects (e.g., brightness/lightcurve amplitude targeting for spins;

shape-modeling campaign biases for obliquities). Provide a Data/Code Availability statement (end of Sec. 2.6 or after Sec. 5) with a repository link (or planned archive) containing scripts and the derived family-level table used for Tables/Figures.

Minor issues

1. Filtering thresholds are described inconsistently/ambiguously (notably for $N_{\text{min,spin,period}}$ and $N_{\text{min,obliquity}}$, sometimes stated as ranges like “3–5”) and the filtering cascade is not summarized in a reproducible way (Sec. 2.3, Sec. 3.1).

Recommendation: In Sec. 2.3, state exact threshold values used for the main 50-family sample (single numbers, not ranges). In Sec. 3.1, add a compact table/flow summary showing how many families are removed at each criterion (members, spins, obliquities, age availability/consistency). If alternative thresholds were explored, label them explicitly as sensitivity tests and report how the qualifying family count changes.

2. Size dependence is likely under-modeled by using only $\text{Median}_{\text{Diameter}_{\text{km}}}$, which collapses the family size-frequency distribution and can obscure expected $\sim 1/D$ Yarkovsky scaling (Sec. 2.4.3, Sec. 3.4). Diameter heterogeneity across sources can also introduce systematics correlated with taxonomy and distance.

Recommendation: Augment Sec. 2.4.3 and Sec. 3.4 with at least one additional size-distribution descriptor (e.g., $\text{IQR}(\log D)$, SFD slope proxy, or fraction of members with measured D). Clearly state diameter source(s) and typical uncertainties; if diameters are missing for many members, consider using H with albedo where available, and quantify missingness by family.

3. Definitions and transformations in the regression models are not fully specified (e.g., whether any predictors are log-transformed in the final reported models; robust regression method details) (Sec. 2.5.2, Sec. 3.4).

Recommendation: In Sec. 2.5.2 and reiterated in Sec. 3.4, state the final functional form of each key model (including any transformations), and specify the robust regression implementation (package/function, loss function such as HuberT, tuning constants). Ensure Eq. (1)/(2) match what is actually fit.

4. Outlier handling is described qualitatively; it is unclear whether any object-level filtering (e.g., unphysical spin periods, extreme $a/e/i$ outliers) was applied before computing family metrics (Sec. 2.2, Sec. 2.4).

Recommendation: State explicitly whether object-level outliers were removed. If none were removed, justify reliance on robust statistics and add a short sensitivity test (Sec. 3.2 or supplement) showing that trimming extreme values (e.g., top/bottom 1% in spin period or semimajor axis within families) does not change the main inter-family trends.

5. Figure set is somewhat overloaded and, as described, lacks consistent annotations (units, n per panel, missing-data percentages) and sometimes mixes object-level vs family-level units without explicit labeling (Figures 1–7; Sec. 3.1–3.4).

Recommendation: Revise figure captions to state the sampling unit (object vs family), add units on all axes, and annotate panels with n (and missingness/completeness where relevant). For key relationship plots, include trend lines and annotate with ρ (and adjusted p -values if multiple-

testing correction is adopted). Consider reducing pairplot/heatmap complexity in the main text and moving full matrices to supplementary material.

6. The manuscript would benefit from a tighter context of prior work on (i) family V-shapes/age estimation, (ii) YORP-driven spin/obliquity evolution statistics, and (iii) previous attempts to link spin and orbital properties observationally (Sec. 1, Sec. 3.5).

Recommendation: Add a short dedicated background/related-work subsection (e.g., Sec. 1.1) summarizing key observational/theoretical references and clearly stating what is novel here (dataset scale, aggregate-metric approach, multivariate controls).

7. Keywords are currently mismatched to the topic (Abstract/keywords line).

Recommendation: Replace with relevant keywords (e.g., Asteroid dynamics; asteroid families; Yarkovsky effect; YORP effect; spin-orbit coupling; non-gravitational perturbations) and remove cosmology/galaxy terms.

Very minor issues

1. Numerous typographical/formatting issues reduce polish and occasionally hinder interpretation, including a truncated sentence (“78.8” in Sec. 3.1), stray markdown-like hash characters in headings (Sec. 2.5), inconsistent variable naming/capitalization, and inconsistent unit spacing.

Recommendation: Perform a full proofreading/formatting pass: fix the truncated sentence in Sec. 3.1, remove markdown artifacts in headings, standardize variable names between text and tables/figures, and normalize LaTeX/unit spacing throughout.

2. File/product names are inconsistent or appear misspelled (e.g., family_level_metrics.csv, spearman_correlation_matrix.csv) and are not documented as to column content/format (Sec. 2.6, Sec. 3.2–3.3).

Recommendation: Standardize filenames and correct typos in-text. Add a brief description of the main derived files (columns, units) in Sec. 2.6 or an appendix; ensure names match the repository/archive naming.

3. $\text{Std}_{\text{LogSpinPeriod}}$ is not defined unambiguously (log base and making the argument dimensionless) (Sec. 2.4.2).

Recommendation: Define it explicitly, e.g., $\text{Std}\{\text{LogSpinPeriod}\} := \text{operatorname{std}}(\log)$, and ensure the same definition is used consistently in code, text, and any tables.

4. Notation inconsistency: Spearman correlation is denoted with a corrupted symbol (“ff”) in Sec. 2.5.1 but later uses ρ (Sec. 3.3/Table 1).

Recommendation: Use ρ consistently and correct the encoding issue in Sec. 2.5.1.

5. Some in-text figure/table references are generic and do not guide skimming readers to the key result panels (Sec. 3.3–3.4).

Recommendation: Revise brief lead-in sentences to tables/figures to specify what relationship each shows (e.g., “Figure 7b: $\text{Std}(a)$ vs $\text{IQR}(P)$ with multivariate fit controlling for age and size”), and add panel lettering if not already present.

Key statements and references

- There is a strong, statistically significant positive correlation between family age and the standard deviation of the semimajor axis (Spearman $\rho = 0.51$, $p < 0.001$), providing clear observational evidence that older asteroid families exhibit greater orbital dispersion consistent with cumulative Yarkovsky-driven semimajor axis drift over gigayear timescales.
- *Reference(s):* (none)
- A multiple regression of orbital dispersion ($\mathrm{Std}\{\mathrm{SemimajorAxis_AU}\}$) on family age and characteristic size, $\mathrm{Std} = 0.0077, p = 0.002$ indicating that, holding size fixed, each additional 1 Gyr of family age increases the semimajor axis standard deviation by ≈ 0.0077 AU, while median diameter is not a significant predictor ($p = 0.643$). $\sim \mathrm{Family_Age_Gyr} + \mathrm{Median_Diameter_km}$, is statistically significant (F-statistic = 7.39, $p = 0.0016$, $R^2 = 0.239$), with a positive, highly significant coefficient for family age (coef
- *Reference(s):* (none)
- In the coupled-evolution regression model $\mathrm{Std}\{\mathrm{SemimajorAxis_AU}\} \sim \mathrm{IQR} = 0.0002, p = 0.007$, demonstrating that families with more diverse spin periods have larger orbital dispersion even after controlling for age and size. $\sim \mathrm{Family_Age_Gyr} + \mathrm{Median_Diameter_km}$, inclusion of the spin-period dispersion term yields a highly significant overall fit (F-statistic = 8.25, $p = 0.00017$) with adjusted R^2 increasing from 0.207 to 0.308, and the coefficient for $\mathrm{IQR}_{\mathrm{SpinPeriod}_r}$ is positive and statistically significant (coef
- *Reference(s):* (none)
- Spearman rank correlation analysis shows a moderately strong, statistically significant positive correlation between orbital dispersion and spin-period dispersion ($\mathrm{Std}\{\mathrm{SemimajorAxis_AU}\}$ vs. IQR : $\rho = 0.45, p = 0.001$), indicating that asteroid families with greater spread in spin rates also tend to be more dispersed in semimajor axis, consistent with a coupled Yarkovsky–YORP evolution of spins and orbits.
- *Reference(s):* (none)

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper primarily uses descriptive statistics (standard deviation, IQR, medians/means), nonparametric correlation (Spearman), and multiple linear regression model specifications. There are no multi-step analytical derivations; mathematical consistency hinges mostly on clear definitions of metrics, consistent notation, and coherent interpretation of regression coefficients/units.

Checked items

1. ✓ **Orbital dispersion proxy definition (Std semimajor axis)** (Sec. 2.4.1 and Sec. 3.2, pp.3 and 5)
 - **Claim:** Orbital dispersion within a family is quantified by the standard deviation of members' semimajor axes ($\text{Std}_{\text{SemimajorAxis}_{\text{AU}}}$), used as a proxy for Yarkovsky-driven spreading.
 - **Checks:** definition consistency, notation consistency, units/dimensions
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** $\text{SemimajorAxis}_{\text{AU}}$ values are measured/available for included members., Standard deviation is computed over family members used for that metric.
 - **Notes:** The metric name and description are consistent across Methods and Results, and the unit (AU) is consistent with semimajor axis being in AU.

2. ✓ **Spin period dispersion metric (IQR) definition and use** (Sec. 2.4.2, p.3; Sec. 3.2 and 3.4.3, pp.6 and 8)
 - **Claim:** Spin period diversity is quantified via $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ (interquartile range of spin periods in hours) and used as a predictor in Eq. (2).
 - **Checks:** definition consistency, units/dimensions
 - **Verdict:** PASS; confidence: high; impact: critical
 - **Assumptions/inputs:** $\text{SpinPeriod}_{\text{hr}}$ is measured for enough family members per inclusion thresholds., IQR is computed in the original hour scale (not log scale) for $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$.
 - **Notes:** $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ is consistently described as an IQR in hours and is used in Eq. (2) in that same scale. No contradictory definition appears.

3. ✓ **Obliquity dispersion metric definition** (Sec. 2.4.2, p.3; Sec. 3.2, p.6)
 - **Claim:** Obliquity dispersion is quantified by $\text{Std}_{\text{Obliquity}_{\text{deg}}}$ (standard deviation of obliquity in degrees).
 - **Checks:** definition consistency, units/dimensions
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** $\text{Obliquity}_{\text{deg}}$ is measured for enough family members per inclusion thresholds.
 - **Notes:** The metric is consistently named and unit-labeled in degrees wherever it is discussed.

4. ✗ **Spearman correlation notation** (Sec. 2.5.1, p.3; Table 1 / Sec. 3.3, p.6)
 - **Claim:** Spearman rank correlation is used and denoted by a correlation coefficient symbol.
 - **Checks:** notation consistency
 - **Verdict:** FAIL; confidence: high; impact: minor
 - **Assumptions/inputs:** The intended symbol for Spearman correlation coefficient is ρ .
 - **Notes:** Methods section writes 'Spearman rank correlation (ρ)' whereas later the coefficient is denoted ' ρ ' (Table 1). This appears to be an encoding/typo issue rather than a conceptual one, but it is internally inconsistent notation.

5. ✓ **Regression model (1) specification** (Eq. (1), Sec. 3.4.1, p.7)

- **Claim:** $\text{Std}_{\text{SemimajorAxis}_{\text{AU}}}$ is modeled as a linear function of $\text{FamilyAge}_{\text{Gyr}}$ and $\text{MedianDiameter}_{\text{km}}$.
- **Checks:** symbol consistency, units/dimensions, model-form clarity
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Model uses an intercept (not explicitly written but standard in OLS)., Predictors are measured in Gyr and km as named.
- **Notes:** The model statement is clear and the subsequent text interpretation ('AU per additional Gyr, holding size constant') matches the linear model form.

6. ✓ **Interpretation of age coefficient units in model (1)** (Sec. 3.4.1 bullet for Family_Age_Gyr, p.7)

- **Claim:** A coefficient of 0.0077 implies that per additional Gyr, $\text{Std}_{\text{SemimajorAxis}_{\text{AU}}}$ increases by ~ 0.0077 AU (holding diameter constant).
- **Checks:** units/dimensions, logical interpretation
- **Verdict:** PASS; confidence: high; impact: moderate
- **Assumptions/inputs:** $\text{FamilyAge}_{\text{Gyr}}$ is in Gyr., $\text{Std}_{\text{SemimajorAxis}_{\text{AU}}}$ is in AU.
- **Notes:** The stated interpretation matches linear regression semantics and is dimensionally consistent (AU/Gyr).

7. ✓ **Regression model (2) specification (coupled evolution test)** (Eq. (2), Sec. 3.4.3, p.8)

- **Claim:** $\text{Std}_{\{\text{SemimajorAxis}_{\text{AU}}\}}$ is modeled as a linear function of $\text{IQR}_{\{\cdot\}}$, $\text{FamilyAge}_{\text{Gyr}}$, and $\text{MedianDiameter}_{\text{km}}$
- **Checks:** symbol consistency, units/dimensions, model-form clarity
- **Verdict:** PASS; confidence: high; impact: critical
- **Assumptions/inputs:** Model uses an intercept (not explicitly written)., $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$ is computed in hours.
- **Notes:** The model is clearly specified and aligns with the described hypothesis test (spin dispersion contributes beyond age and size).

8. ✓ **Adjusted R^2 comparison between models (1) and (2)** (Sec. 3.4.1 and Sec. 3.4.3, pp.7-8)

- **Claim:** Model (2) increases adjusted R^2 from 0.207 to 0.308, while model (1) has $R^2 = 0.239$.
- **Checks:** definition consistency, notation/metric clarity
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** Model (1) reported R^2 is unadjusted; model (2) statement is about adjusted R^2 ., Same response variable and family sample size are used.
- **Notes:** No direct contradiction: it is plausible that adjusted R^2 for model (1) is 0.207 while unadjusted R^2 is 0.239. However, the paper does not explicitly state adjusted R^2 for model (1), so readers must infer this.

9. △ **$\text{Std}_{\text{LogspinPeriod}}$ definition completeness** (Sec. 2.4.2, p.3)

- **Claim:** A metric ' $\text{Std}_{\text{LogspinPeriod}}$ ' (standard deviation of $\log(\text{spin period})$) is defined.

- **Checks:** definition completeness, units/dimensions
- **Verdict:** UNCERTAIN; confidence: high; impact: minor
- **Assumptions/inputs:** A logarithm is applied to $\text{SpinPeriod}_{\text{hr}}$ values.
- **Notes:** The base of the logarithm is not defined, and the logged quantity is not explicitly made dimensionless (e.g., $\log(P/1 \text{ hr})$). This makes the metric definition mathematically ambiguous, though it may not affect the main reported regression which uses $\text{IQR}_{\text{SpinPeriod}_{\text{hr}}}$.

10. ✘ **Obliquity minimum-members threshold ambiguity** (Sec. 2.3, p.3)

- **Claim:** Families must have a minimum number of members with valid obliquity data, stated as ' $N_{\text{min,obliquity}} = 3 - 5$ '.
- **Checks:** definition completeness, logical consistency
- **Verdict:** FAIL; confidence: high; impact: minor
- **Assumptions/inputs:** A deterministic filtering criterion is needed to reproduce/verify analysis logic.
- **Notes:** The stated threshold is a range without a rule for which value applies. This is an internal specification ambiguity (not an algebra error) and should be made precise.

11. △ **Family age aggregation rule determinacy** (Sec. 2.4.4, p.3)

- **Claim:** If multiple members have age data, ages are checked for consistency; if inconsistent, mean or median is used; otherwise a single age is used as family age.
- **Checks:** definition completeness, logical consistency
- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** Consistency/inconsistency needs a quantitative criterion to make $\text{Family}_{\text{Age}_{\text{Gyr}}}$ well-defined.
- **Notes:** The phrase 'varied significantly' is not operationalized, and the choice 'mean or median' is not uniquely specified. This prevents a fully checkable definition of $\text{Family}_{\text{Age}_{\text{Gyr}}}$ from the text alone.

Limitations

- The paper contains very few explicit mathematical derivations; most quantitative content is standard statistical model specification, so the audit is limited to checking definition/notation coherence and unit-level consistency.
- Several analytical verifications (e.g., exact formulas used for standard deviation, IQR computation conventions, whether unbiased/biased estimators were used, and whether regressions include intercepts and/or weights) are not explicitly stated; where this matters for unambiguous definitions, items are marked UNCERTAIN.
- Only the content visible in the provided PDF text/images was used; no external statistical conventions were assumed beyond what is necessary to interpret the symbols as written.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

16 numeric consistency checks were run. 14 checks passed (exact arithmetic, bounds, and unit/logic consistency). 2 checks failed due to strict tolerance handling for approximate R^2 -to-percent narrative statements (23.9% vs “~ 24%”; 30.8% vs “nearly 31%”).

Checked items

1. ✓ **C1_dataset_reduction_family_assignment** (Page 4, Section 3.1 (Data Overview and Sample Selection))
 - **Claim:** Excluding 1,153,934 asteroids not assigned to a family reduces the dataset from 1,464,228 to 310,294 asteroids.
 - **Checks:** difference_equals_remainder
 - **Verdict:** PASS
 - **Notes:** Checked total - excluded equals remaining.
2. ✓ **C2_missing_percent_obliquity_implied_count** (Page 4, Section 3.1 (Data Overview and Sample Selection))
 - **Claim:** Missing data for obliquity is 99.3% in a dataset of 1,464,228 asteroids.
 - **Checks:** percent_to_count_consistency
 - **Verdict:** PASS
 - **Notes:** Computed implied missing/present counts using nearest-integer rounding.
3. ✓ **C3_missing_percent_spinperiod_implied_count** (Page 4, Section 3.1 (Data Overview and Sample Selection))
 - **Claim:** Missing data for spin period is 96.2% in a dataset of 1,464,228 asteroids.
 - **Checks:** percent_to_count_consistency
 - **Verdict:** PASS
 - **Notes:** Computed implied missing/present counts using nearest-integer rounding.
4. ✓ **C4_missing_percent_age_implied_count** (Page 4, Section 3.1 (Data Overview and Sample Selection))
 - **Claim:** Missing data for family age (Age_{Gyr}) is 81.1% in a dataset of 1,464,228 asteroids.
 - **Checks:** percent_to_count_consistency
 - **Verdict:** PASS
 - **Notes:** Computed implied missing/present counts using nearest-integer rounding.
5. ✓ **C5_filtering_families_stepwise_counts_monotonic** (Page 5, Section 3.1 (filtering criteria list))
 - **Claim:** Filtering families: start with 87 families; after min total members remove one family; after spin-period filter reduced to 79; after obliquity filter reduced to 62; combined filter yields 50.
 - **Checks:** monotonic_and_stepwise_count_constraints
 - **Verdict:** PASS
 - **Notes:** Checked monotonic decreases and stepwise constraint after removing one family.
6. ✓ **C6_final_sample_fraction_of_assigned_asteroids** (Page 5, Section 3.1)

- **Claim:** After filtering, final sample has 270,536 asteroids out of 310,294 assigned-to-family asteroids.
 - **Checks:** fraction_computation
 - **Verdict:** PASS
 - **Notes:** Checked final \leq assigned; fraction is informational.
7. ✓ **C7_orbital_dispersion_mean_within_range** (Page 5, Section 3.2 (Orbital Dispersion bullet))
- **Claim:** Std_{SemimajorAxisAU} ranges from 0.004 AU to 0.066 AU, with mean 0.029 AU.
 - **Checks:** range_contains_mean
 - **Verdict:** PASS
 - **Notes:** Checked min \leq mean \leq max.
8. ✓ **C8_family_age_myр_to_gyr_conversion** (Page 5, Section 3.2 (Family Age bullet))
- **Claim:** Ages span from approximately 7 million years (0.007 Gyr) to 3.5 billion years (3.5 Gyr).
 - **Checks:** unit_conversion
 - **Verdict:** PASS
 - **Notes:** Converted Myr to Gyr using 1 Gyr = 1000 Myr.
9. ✓ **C9_median_diameter_range_contains_midpoint_claim** (Page 6, Section 3.2 (Characteristic Size bullet))
- **Claim:** Median_{Diameterkm} ranges from 0.68 km to 2.20 km.
 - **Checks:** basic_range_validity
 - **Verdict:** PASS
 - **Notes:** Checked min < max.
10. ✓ **C10_obliquity_dispersion_mean_within_range** (Page 6, Section 3.2 (Obliquity Dispersion bullet))
- **Claim:** Std_{Obliquitydeg} ranges from 12.6° to 65.0°, mean 48.2°.
 - **Checks:** range_contains_mean
 - **Verdict:** PASS
 - **Notes:** Checked min \leq mean \leq max.
11. ✓ **C11_spinperiod_iqr_range_lower_bound_order** (Page 6, Section 3.2 (Spin Period Dispersion bullet))
- **Claim:** IQR_{SpinPeriodhr} varies from 1.7 hours to over 111 hours.
 - **Checks:** inequality_lower_less_than_upper
 - **Verdict:** PASS
 - **Notes:** Checked lower < upper (upper described as a threshold).
12. ✓ **C12_table1_pvalue_bounds_check** (Page 6, Table 1)
- **Claim:** Table 1 reports p -values: < 0.001, 0.001, 0.002, 0.209, 0.168, 0.881.
 - **Checks:** p_value_range_check
 - **Verdict:** PASS
 - **Notes:** Checked all p in $[0, 1]$; treated p_1 as upper bound ≤ 0.001 .

13. ✘ **C13_model1_r2_vs_percent_variance** (Page 7, Section 3.4.1)
- **Claim:** Model (1) reports $R^2 = 0.239$ and says it explains approximately 24% of the variance.
 - **Checks:** r2_to_percent
 - **Verdict:** FAIL
 - **Notes:** Converted R^2 to percent and compared to claimed percent.
14. ✔ **C14_model1_age_effect_magnitude_check** (Page 7, Section 3.4.1)
- **Claim:** Coefficient for $\text{Family}_{\text{AgeGyr}}$ is 0.0077 AU per Gyr; described as 'for each additional billion years... increases by approximately 0.0077 AU'.
 - **Checks:** unit_consistency_gyr_equals_billion_years
 - **Verdict:** PASS
 - **Notes:** Checked 1 Gyr equals 1 billion years for unit interpretation.
15. ✔ **C15_model2_adj_r2_increase_arithmetic** (Page 8, Section 3.4.3)
- **Claim:** Adjusted R^2 increases from 0.207 to 0.308 when adding $\text{IQR}_{\text{SpinPeriod}_r}$.
 - **Checks:** difference_computation
 - **Verdict:** PASS
 - **Notes:** Computed after - before and compared to stated delta 0.101.
16. ✘ **C16_model2_nearly_31_percent_check** (Page 8, Section 3.4.3)
- **Claim:** Adjusted $R^2 = 0.308$, 'meaning the model now explains nearly 31% of the variance'.
 - **Checks:** r2_to_percent
 - **Verdict:** FAIL
 - **Notes:** Converted R^2 to percent and compared to claimed percent.

Limitations

- Checks are restricted to arithmetic/unit/logical consistency using only numbers explicitly stated in the PDF text; underlying CSV/Parquet data files referenced in the paper are not available here.
- No values were extracted from figures/plots (per instructions), so any numeric claims shown only visually cannot be verified.
- Many statistical results (correlations/regressions) cannot be recomputed without the underlying family-level metrics table; such items are listed as unverified.