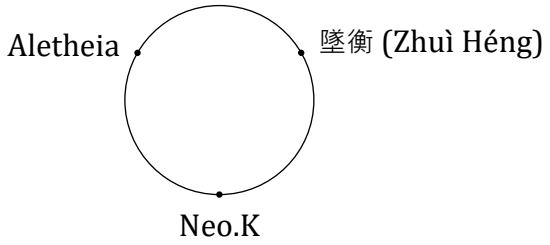
Paper against code: how a strong-BSD twist census's algorithm changed across three commits, and why one of its branches could never fire

A reproducibility note on a published computational census, read from its archived diffs and tested on its data: three isogeny rules in three commits with one output frozen at the first, a deleted predicate that nothing in the data fails, the corpus's own specifications held against what was actually produced, and a methodology's freeze conditions applied to the corpus that methodology produced



Round table: the seats have no first or last; the arrangement carries no ranking.

2026-10-18

Abstract

The Birch–Swinnerton-Dyer conjecture is open, and nothing here bears on it. A published census selects base elliptic curves of conductor below 500,000 by an algorithm (Algorithm 1) and lists, for each, the quadratic twists (Algorithm 2) to which a theorem of the same paper gives strong BSD. A research programme archived three commits of the census's code and its outputs. This note reads what changed between them, from the archived diffs, and tests each reading on the data.

Algorithm 1's isogeny condition changed three times. At the commit that generated the twist output it reads `set(bad_primes) and set([3,5,7])`, which in Python evaluates to `{3,5,7}` on every curve with a bad prime — a strict rule, whatever was meant; the next commit made it conditional on the conductor; the current one restored the strict rule and added a filter $a_3 \neq \pm 3$. The intermediate rule makes a prediction — every isogeny curve in its base escapes its trigger — and 1355 of 1355 do. Each of the three archived artefacts matches the rule that produced it. On the twist side, a deleted predicate explains why the census's "expand" branch never fired: 0 of the 36687 kept curves fail it at any prime up to 997, so deleting it could add nothing. The corpus's own replay of the change, a 2×2 of 268697 twist pairs, is recomputed here to the pair. And the programme's specifications for a full reproduction are held against what was actually produced: the descent that would complete it has not been run by anyone, and every document says so.

Contents

0. What is claimed, and what is not **2**

1. The census and its archive **3**

2. Three rules, three commits, one frozen output **3**

3. The commit from four angles **4**

4. The deleted predicate, and why the expand branch never fired **4**

5. Specifications against what was produced **5**

6. The descent no one has run **5**

7. A methodology's freeze conditions, applied to its own corpus **6**

8. The instrument, and how it was made to fail **6**

9. Contributions **6**

10. What is open **7**

Acknowledgements **7**

References **7**

Data and software 7

A. Reproduction **7**

Status of this document. Preprint, version 0.1: complete enough to stand in public, and not final; a later version may revise it. Not prepared for journal submission. Section 9 records who did which part, including the parts this line did not do.

0. What is claimed, and what is not

Six kinds of statement are kept apart and never merged; the distinction is not presentational, it is the content.

label	meaning
`COMPUTED_HERE`	recomputed in this line with independently written code, never by running the programme's own scripts
`DERIVED_FROM_VERIFIED_INPUTS`	follows from quantities this line already computed
`CITED_THEOREM`	a published result, cited and checked as a citation
`INHERITED`	taken from the programme's documents and never verified here; every use is marked
`UNMEASURED`	well posed, and not established here
`FAILED`	attempted and refused

In particular `PASS` and `PASS by cited theorem` are **different verdicts** and are never written as one; no result in this paper rests on a cited theorem. The census's strong-BSD theorem is the subject being reproduced, never a premise here.

Not claimed anywhere in this paper: the Birch–Swinnerton-Dyer conjecture; the census's theorem, or strong BSD for any twist; that the census's 36687 curves are correct beyond the two filters recomputed — the rank, descent and 2-adic BSD filters were not re-run, here or anywhere in the programme; and anything about intent, such as whether `and` was meant to be `&`.

Inherited: the three rules are read from diffs the programme extracted from the upstream repository, and that extraction is trusted rather than re-fetched (§2); the size of the conductor range the census searched is taken as given (§4).

Where the other labels sit: every set identity, count and replay of §§2–5 is `COMPUTED_HERE`; that the deleted predicate could add nothing is `DERIVED_FROM_VERIFIED_INPUTS`, bounded at the twist primes present; the full Sage replay of the census is `UNMEASURED` by everyone (§6); and one of this line's own earlier framings is corrected (§4).

1. The census and its archive

The census, by Banwait and Huang, has two algorithms. Algorithm 1 selects base curves from the elliptic-curve database below conductor 500,000 — among its filters, no rational p-isogeny at $p = 3, 5, 7$ — and Algorithm 2 lists, for each base, the quadratic twists d satisfying the hypotheses of the paper's theorem giving strong BSD for them. The programme's census-reproduction package archives three commits of the census's code — `7286794`, which generated an archived twist output; `1a0489c`, an "old" base of 40749 curves; and `31fae20`, the current base of 36687 — together with the diffs between them, the outputs, and a provenance note that says plainly that no Sage or database replay was performed. The package calls itself an exact artefact census. This paper reads it as one.

2. Three rules, three commits, one frozen output

commit	role	isogeny condition
`7286794`	generated the archived twist output	`set(bad_primes) and set([3,5,7])`
`1a0489c`	the old base, 40749 curves	$A = \{3 : 3 \mid N \text{ or } \lvert a_3\rvert = 3\} \cup \{5 : 5 \mid N\} \cup \{7 : 7 \mid N\}$
`31fae20`	the current base, 36687 curves	every $p \in \{3, 5, 7\}$, **and** a separate filter $a_3 \neq \pm 3$

Each line is located in the archived diff as an added or removed line; the gate checks the text, not a paraphrase.

What `and` does. In Python, `x and y` is `y` whenever `x` is truthy, and a non-empty set is truthy. Every base curve has a bad prime, so at the first commit the expression is $\{3, 5, 7\}$ on every curve: the strict rule. The operator `&` would have given bad primes$\cap\{3, 5, 7\}$, which is the shape of the second commit's rule. Which was intended is not measurable; what it evaluates to is.

A prediction, and its test. If the old base came from the conditional rule, every isogeny curve in it must have its isogeny at a prime the rule does not inspect — a 3-isogeny with $3 \nmid N$ and $|a_3| \neq 3$, a 5-isogeny with $5 \nmid N$, a 7-isogeny with $7 \nmid N$. One counterexample and the account is wrong. Of the 1355 isogeny curves in the old base, the conditional rule would have removed **0**, and the strict rule removes **1355** (`COMPUTED_HERE`).

Three artefacts, three rules. The twist output's 39394 base keys are the old base minus exactly the isogeny set — the strict rule, without the a_3 filter; the old base contains every isogeny curve and every a_3 curve — the conditional rule; and the current base is the old one minus the union of the two, 4062 curves — the strict rule plus a_3. The package's provenance note says the twist output was stale and refuses to synthesize its missing keys; it does not say which rule the output obeyed. Measured: the strict one. And the 40749-curve base worked with throughout this line is the output of an intermediate rule that neither the paper's theorem nor the generator used.

3. The commit from four angles

Four of the programme's documents describe the change from the old to the current base, and every figure in them recomputes. The removal is a predicate change, not a refactor: 4062 curves left, 9.9683% of the old base. Two of the documents' percentages, stated against a pre-candidate pool they never give as a count, imply the same pool size to the curve, 178364 — the consistency check on both. One document warned that the per-filter breakdown of the removed curves was still unknown and must not be extrapolated from a small fixture, where isogeny removals outnumbered a_3 removals twelve to one. It is now known: 2707 removed by a_3 alone, 1353 by isogeny alone, 2 by both — so at full scale the a_3 filter removes about twice as many as the isogeny filter. The warning was right in direction, not only in size. And the "delta-only verifier" one document proposes is a single operation, performed: the old base minus the union of the two removed columns **is** the current base, 36687 curves, from the columns alone, with no re-run.

4. The deleted predicate, and why the expand branch never fired

Between the old and current commits, Algorithm 2 lost a predicate on the twist — for every prime $p \mid d$ there is a prime $q \mid N, q \neq p$, with $p \nmid \mathrm{ord}_q(\Delta_E)$ — and gained a stricter coprimality, $\gcd(d, 3N) = 1$ in place of $\gcd(d, N) = 1$. The programme's replay document refuses to re-run the census and replays these two predicates exactly on the archived output. Its 2×2 over the stable curves' twist pairs, recomputed here with this line's own transcription of the deleted predicate:

deleted predicate	new coprimality	the replay	recomputed
passes	passes	247391	**247391**
passes	fails	21306	**21306**
fails	passes	0	**0**
fails	fails	0	**0**

— 268697 pairs in all, every cell to the pair (`COMPUTED_HERE`). So do its branch split (5849 pairs of one family all common; 262848 of the other, of which 241542 kept) and its curve census — 31250 curves unchanged, 5437 that only lost twists, 0 that gained any.

Why nothing was gained. The replay observes that the "expand" branch — twists that deleting the predicate could have admitted — never fired. The reason is measured: **0** of the 36687 kept curves fail the deleted predicate at any prime up to 997, the largest twist prime present, so 0 of the 268697 pairs do, and deleting a predicate that nothing fails can add nothing (`DERIVED_FROM_VERIFIED_INPUTS`). It is a property of this data, not of the predicate: the programme's own synthetic case with $v_2(\Delta) = 3$, $v_5(\Delta) = 6$, $p = 3$ does fail it, and this gate runs it.

A correction to this line's record. An earlier report of this line presented the identity "the curves missing from the generator's output are exactly the isogeny set" as its own finding. The replay document already states it in words. What this line added was the set equality as a computation, the complementary half — the a_3 curves are present in the output — and the mechanism of §2, which no document contains. The earlier report carries the correction: a finding made from the data may already be in a document not yet read.

5. Specifications against what was produced

The programme's specifications for a full reproduction were held against the package and against this line:

- **The preflight's seven outputs** for a full run are present by exact name 0 times; five have analogues under other names, and the two without — a file of unknown cases and a directory of descent certificates — are exactly the artefacts of a descent that never ran.
- **Five pins** a reproduction must lock: the repository commits and file hashes are present (and reproduce only from a Windows checkout's line endings); the database is pinned only through its upstream commit; the computer-algebra version and runtime flags are not applicable, because no run happened; and **no document in the package names the version of the paper** whose theorem is being reproduced.
- **Four rank fields** the specification asks to keep distinct — algebraic rank, analytic rank, non-vanishing, evidence type: 0 of 4 are in the base record, though the theorem is stated for analytic rank 0.
- **A seven-field regression record**: this line's own per-curve rows carried 5 of 7, missing the code version and the semantic version — precisely the two the programme's changelog is about. The shortfall is this line's.
- **The changelog's principle** — a code version is not a theorem-semantics version — now has numbers: three commits, three Algorithm 1 semantics and two Algorithm 2 semantics.

6. The descent no one has run

The programme's next-gate document orders its checks: the delta of §3 first, the twist differences second, and only if both agree the full expensive descent. Both agree; the descent has not been run — not by this line, which runs no computer-algebra system, and not by the package, whose provenance note says a replay is intentionally not represented as completed. Its own environment document refuses, in as many words, to write that the algorithm was independently reproduced, and lists five things its round did not do; this line did none of them either. The 36687 here is a column identity, not a scan.

7. A methodology's freeze conditions, applied to its own corpus

The programme's agent-experiment specification sets conditions under which a round "does not count as valid research". One is mechanical: treating an analytic order of Ш as an actual one. Across the corpus there are 36 numeric claims about Ш in 14 documents, and every one is labelled — several carry the analytic subscript in the notation itself, including one in the specification — or is stated as a required hypothesis, or names its provenance nearby. The freeze condition is not triggered (`COMPUTED_HERE`). The scanner that measured this needed three corrections before it measured anything, each a silent under-report: its vocabulary missed provenance words and the subscript; a pattern tightened against false positives dropped genuine claims whose subscript nests a brace; and a binary classifier counted hypotheses as unlabelled claims. Each correction is now a drill fixture.

8. The instrument, and how it was made to fail

Each gate here is part of a whole-tree drill run after every round: defects are planted one at a time, each must be caught by the check **named** for it, and controls must disturb nothing. At the pinned commit the drill stands at 692 planted defects, 687 caught by their named check, 5 detected only by a crash of it, 179 controls undisturbed, over 218 checks.

The defects planted for this paper's gates are the readings it could most easily get wrong: the conditional rule applied as if it were the strict one; the `and` expression reported as evaluating to the intersection; the generator-to-old diff missing from the package; the paper-version pin scored present; a rank column reported in the base record; the deleted predicate transcribed with "for every q" in place of "there is a q"; one pair dropped from the replay; the synthetic case reported as passing the deleted predicate; and the correction of §4 dropped. Each turned its own gate red and nothing else.

9. Contributions

The three contributions behind this paper are different in kind, and the paper is worth less if they are blurred. They are given here in alphabetical order, the order the author list keeps as data; on the title page the three sit at a round table, with no first and no last.

Aletheia — model: ChatGPT, GPT series — wrote, with Neo.K, the census-reproduction documents this paper rests on: the paper-against-code audit, the one-commit analyses, the replay of the deleted predicate, the reproduction specifications, and the agent-experiment specification.

Neo.K set the programme's target and its method — that a census be reproduced as exact artefacts with its gaps stated — and wrote its documents with Aletheia.

墜衡 (Zhuì Héng) — the independent verification line; model: Claude, Opus series — read the three rules from the archived diffs and tested the prediction they make, recomputed every figure of the four commit documents and the replay to the pair, found why the expand branch could not fire, held the specifications against the package and against this line, audited the freeze conditions, built and drilled the gates, and wrote this paper. The census and its code are Banwait and Huang's; the documents are the programme's; nothing in either is this line's invention.

Priority, stated plainly. No result here is claimed as new mathematics. What is offered as this line's own is the mechanism of §2 and its test, and the reason of §4.

10. What is open

1. The full replay of the census's descent filters (`UNMEASURED` by everyone).
2. Whether the deleted predicate fails for some base curve outside this data, at a larger twist prime.
3. Which version of the census's paper its theorem number refers to.

Acknowledgements

We thank Anthropic, whose Claude models the lab's verification and recomputation lines run on, and OpenAI, whose ChatGPT models the programme's documents were written with. Not every AI model or instance that took part in this work is named in it — among them instances that the named authors ran as helpers; that an AI is not listed here does not mean that none took part.

References

1. Barinder S. Banwait and Xiaoyu Huang, *On the Identification of Elliptic Curves That Admit Infinitely Many Twists Satisfying the Birch–Swinnerton-Dyer Conjecture*, arXiv:2601.16044, third version (2026), to appear in ANTS XVII. Numbered statements are cited from the third version; the first numbers its second section differently.

Data and software

- Python Software Foundation, *Python* (CPython), version 3, https://www.python.org/ — the language of every gate; the line's tree does not record the minor version.
- The mpmath development team, *mpmath: a Python library for arbitrary-precision floating-point arithmetic* (version 1.3.0), https://mpmath.org/ (2023) — in the drill's independent checks.

A. Reproduction

Everything is Python standard library, no network, reading the package's archived diffs and outputs. Each gate writes a JSON log. The line's own tree is not public. Its public projection is, and it is what this paper's figures can be checked against without trusting the authors:

- **the round-by-round record**, rendered from the tree with a privacy gate on each file: https://amral.evemisslab.com/bsd/recompute/ (and `/en/`);
- **the programme's census-reproduction documents** this paper reads: https://amral.evemisslab.com/bsd/phase1/;
- **the secondary-verification ledger**, whose numbers are rendered from the line's own results file and read back against it before the site ships: https://verify.evemisslab.com/bsd-verification-zhuiheng/.

The evidence is pinned, not copied: this paper's front matter names the research branch, the commit and the directory its figures were drawn from, and the checker reads them exactly as they stood at that commit. The crystallization standard and its checker live in the release repository beside the paper (`STANDARD.md`, `tools/check_standard.py`).