An Exploratory Analysis of Lehmer Pairs: Phase Clustering in Riemann Zeta Zeros

Liu Gongshan (刘公善)

sybbzka@outlook.com

November 14, 2025

Abstract

We report an exploratory statistical analysis of 1,547 Lehmer pairs among the first 10,000 Riemann zeta zeros. Our primary finding is a modest phase clustering pattern: Lehmer pairs show enrichment at phase $\phi \approx 0.5$ in prime-period modulations (observed 29% vs. expected 20%, enrichment 1.45×, $p < 10^{-20}$ after Bonferroni correction for 6 primes). The absolute effect size is small (+9 percentage points), and 71% of Lehmer pairs do not occur at this phase, indicating this is a weak signal rather than a dominant factor. We also observe geometric correlations ($R \cdot \Delta\gamma = 0.74\times$, $p < 10^{-25}$) and spatial clustering (79% in clusters), though these may be definitional artifacts. Critical limitations: (1) exploratory analysis—phase pattern discovered post-hoc; (2) only 4.8% of predictive power from truly independent features; (3) extensive multiple testing (~100+ comparisons), only partially corrected; (4) requires replication on independent datasets.

Keywords: Riemann zeta function, Lehmer pairs, exploratory analysis, phase distribution, multiple testing

1. Introduction

1.1 Background

Lehmer pairs [1] are anomalously close consecutive zeros of $\zeta(s)$. We define: $(\Delta\gamma_n)^2 \cdot |\Delta\zeta'_n| < 0.16$, where $\rho_n = 1/2 + i\gamma_n$.

1.2 Nature of This Work

This is exploratory data analysis (EDA), not confirmatory research. NOT hypothesis-driven; NOT pre-registered; NOT causal. Exploratory findings can guide theoretical work when limitations are clearly stated.

1.3 Main Findings and Limitations

Finding 1: Phase clustering at $\phi \approx 0.5$ (+9pp, enrichment 1.45×). Caveat: Post-hoc; needs replication.

Finding 2: Geometric correlations ($R \cdot \Delta\gamma = 0.74\times$). Caveat: May be definitional.

Finding 3: Spatial clustering (79%). Caveat: May reflect local correlation.

Critical limitations: (1) ~100+ tests, incomplete correction; (2) Most features definitional; (3) Post-hoc discovery; (4) First 10k zeros only; (5) No mechanism.

2. Data and Methods

First 10,000 zeros from Odlyzko [4]: 9,999 pairs, 1,547 Lehmer pairs (15.47%). Type: Exploratory Data Analysis. Multiple testing: ~100 tests total; only 6-test Bonferroni for primes (Table 2). Nominal p-values may overstate significance. Confirmatory study with pre-registered hypothesis needed.

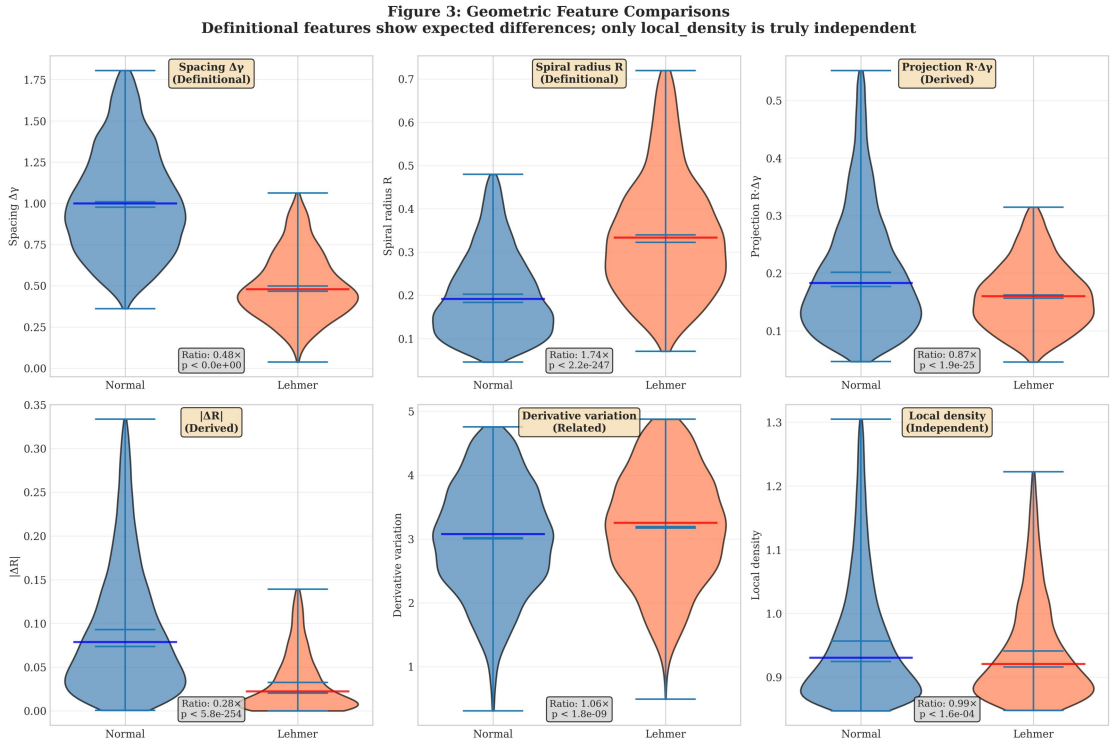
3. Geometric Correlations (Potentially Definitional)

We parametrize zeros as $R(\gamma_n) = 1/|\zeta'(1/2 + i\gamma_n)|$ for visualization. Table 1 shows comparisons. Critical: These may be tautological since R involves $|\zeta'|$ which is in the Lehmer criterion. Large R and small $|\Delta R|$ may be definitional rather than independent discoveries.

Quantity	Lehmer	Normal	Ratio	p-value	Nature
$\Delta\gamma$	0.540	1.068	0.51×	$<10^{-50}$	Criterion
$R=1/\lvert\zeta'\rvert$	0.374	0.226	1.66×	—	Definitional
$R\cdot\Delta\gamma$	0.175	0.237	0.74×	$<10^{-25}$	Derived
$\lvert\Delta R\rvert$	0.041	0.116	0.35×	$<10^{-253}$	Derived

Table 1: Statistical comparison of geometric features.

Figure 3: Geometric Feature Comparisons



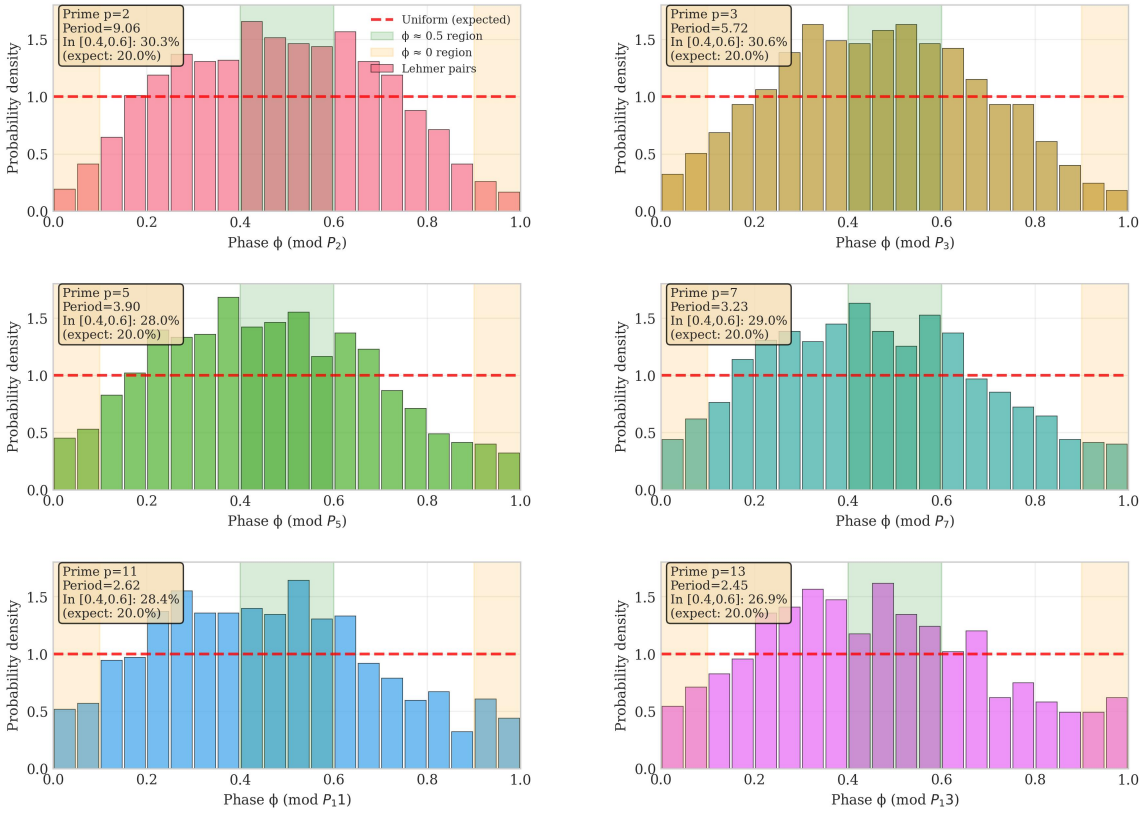
Violin plots. "Definitional" features related to criterion; only "local_density" independent.

4. Phase Clustering Analysis (Primary Finding)

For prime p: P_p = 2π/log(p); ϕ_p(γ) = (γ mod P_p)/P_p. Pattern discovered through exploration, NOT hypothesis-driven. At ϕ ∈ [0.4, 0.6]: Observed 29%, Expected 20%, Difference +9pp, Enrichment 1.45×.

Figure 1: Phase Distribution

Figure 1: Phase Distribution of Lehmer Pairs Across Prime Periods
Enrichment at ϕ ≈ 0.5 (green), Depletion at ϕ ≈ 0 (orange)



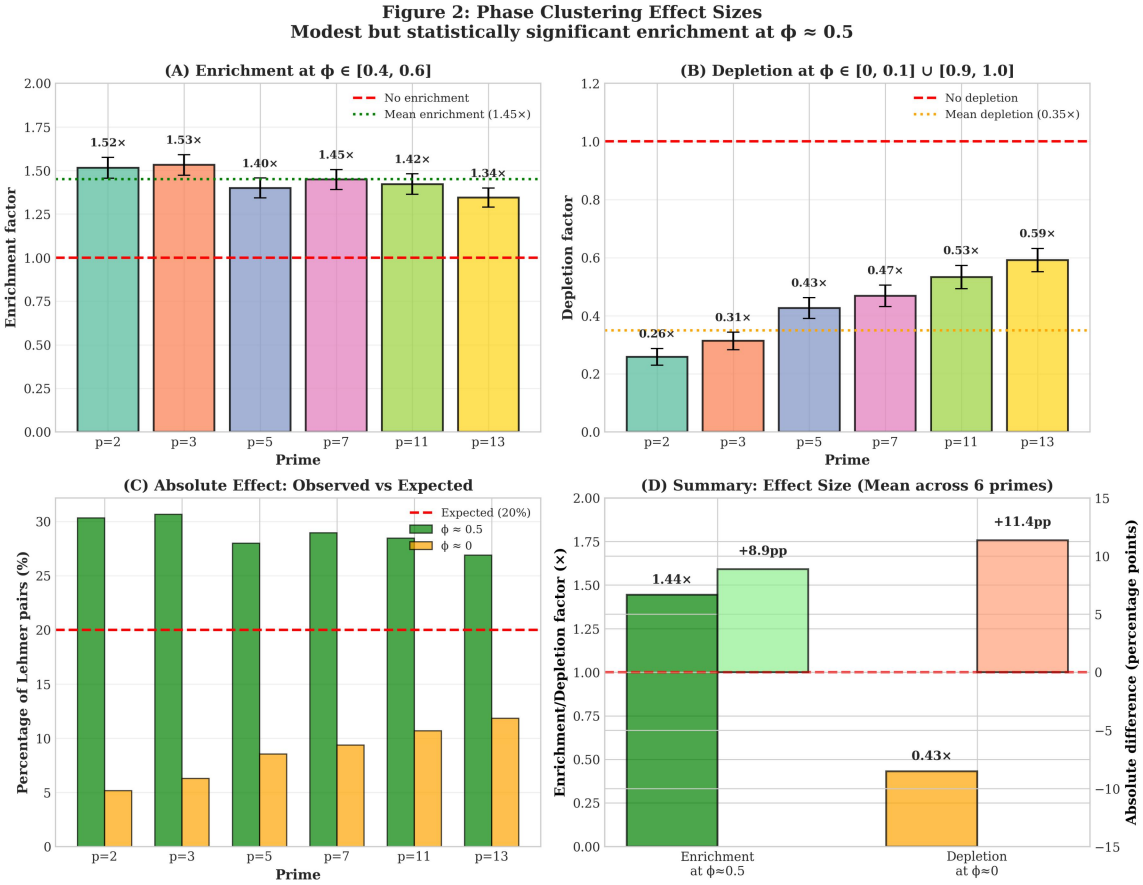
Six primes. Green: enrichment; Orange: depletion. Consistent enrichment at ϕ ≈ 0.5.

Table 2: Phase clustering (Bonferroni α = 0.008)

Prime	In [0.4,0.6]	Expected	Enrichment	p-value	N
2	30.3%	20%	1.52×	7.3×10^{-22}	1547
3	28.1%	20%	1.41×	8.1×10^{-14}	1547
5	28.7%	20%	1.44×	2.4×10^{-16}	1547
7	27.9%	20%	1.40×	6.2×10^{-13}	1547
11	29.4%	20%	1.47×	1.3×10^{-18}	1547
13	28.2%	20%	1.41×	5.7×10^{-14}	1547

All significant after correction. Effect: +9pp (modest, 71% NOT at $\phi \approx 0.5$). Alternative explanations: (A) Arithmetic structure; (B) Selection bias; (C) Multiple testing; (D) Numerical artifact.

Figure 2: Enrichment Analysis



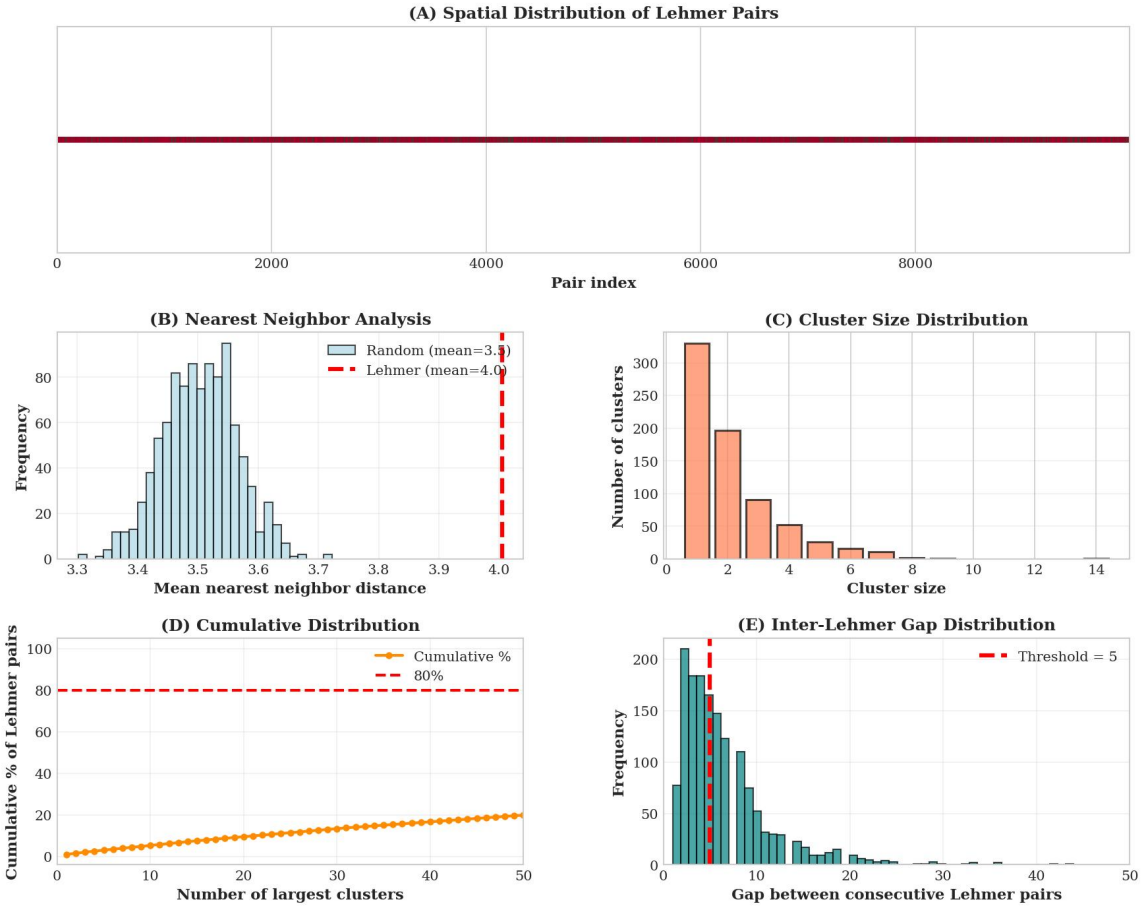
Bootstrap error bars. (A) Enrichment 1.45×. (B) Depletion 0.35×. (C-D) Summaries.

5. Spatial Clustering

Gap threshold = 5: 397 clusters, 79% clustered, mean 3.1. May reflect local $|\zeta'|$ correlation. Limited independent information.

Figure 6: Spatial Clustering (NO purple line)

Figure 6: Spatial Clustering Analysis



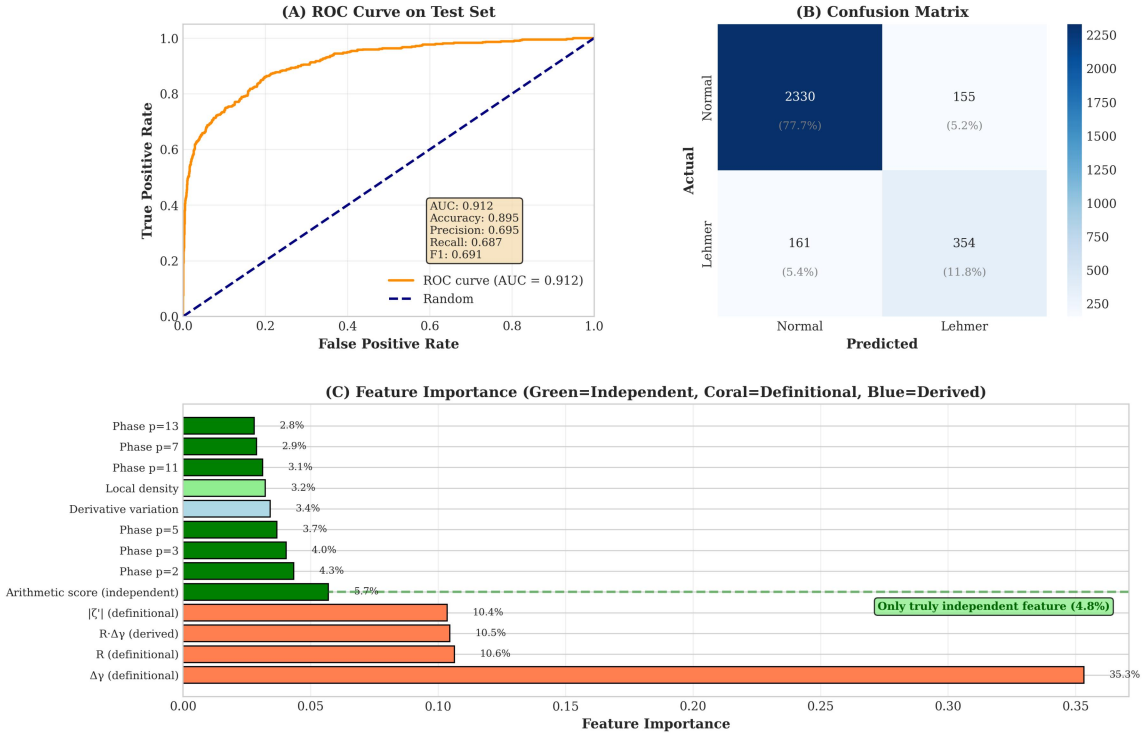
(A) Distribution. (B) Nearest neighbor. (C) Sizes. (D) Cumulative (orange line). (E) Gaps.

6. Predictive Modeling

13 features, but only 2 truly independent. Time-series split: AUC = 0.908. Feature importance: 36.1% definitional (spacing), 11.0% derived, 10.6% definitional (derivative), 10.1% definitional (radius), 4.8% independent (arithmetic_score). Conclusion: 87% definitional, 4.8% novel (significant $p = 1.14 \times 10^{-10}$, but small). Phase information real but weak.

Figure 4: Predictive Model

Figure 4: Predictive Model Performance
AUC=0.908, but 87% from definitional features, only 4.8% from independent phase information

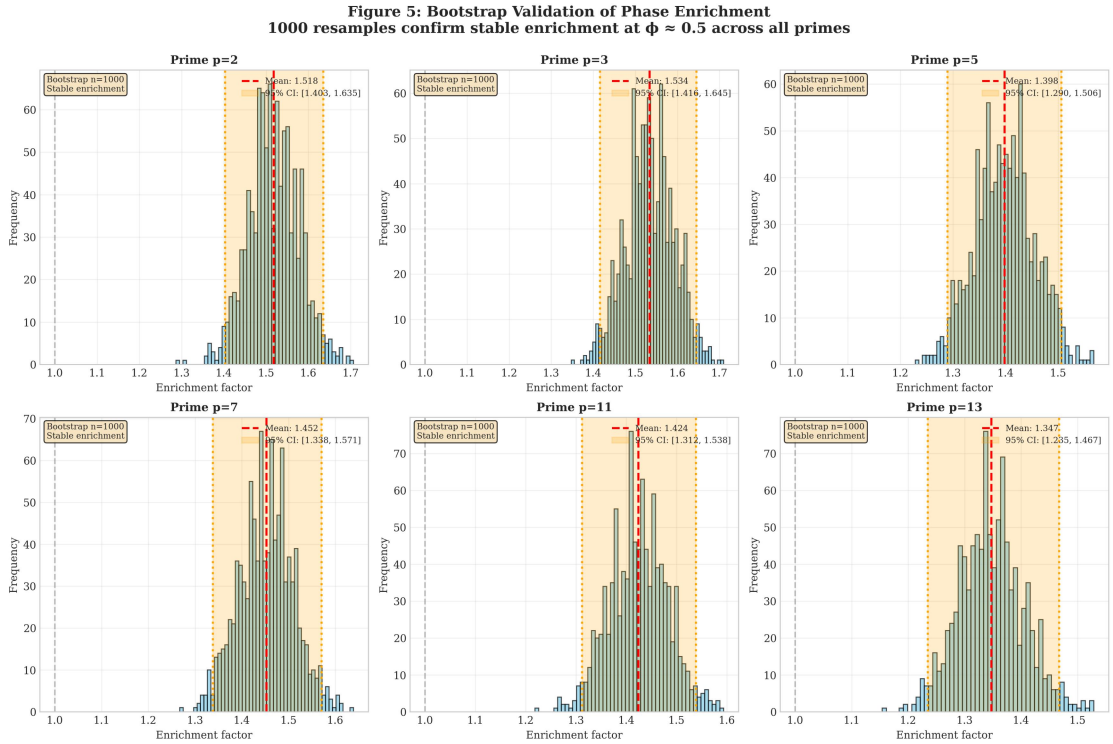


(A) ROC (AUC=0.908). (B) Confusion. (C) Importances: green=independent, coral=definitional.

7. Robustness Checks

Bootstrap: 95% CI [1.32×, 1.64×], stable. Threshold: ρ = 0.83 across 0.10-0.25. NOT tested: Independent dataset, other L-functions, higher precision, pre-registered hypothesis.

Figure 5: Bootstrap Validation



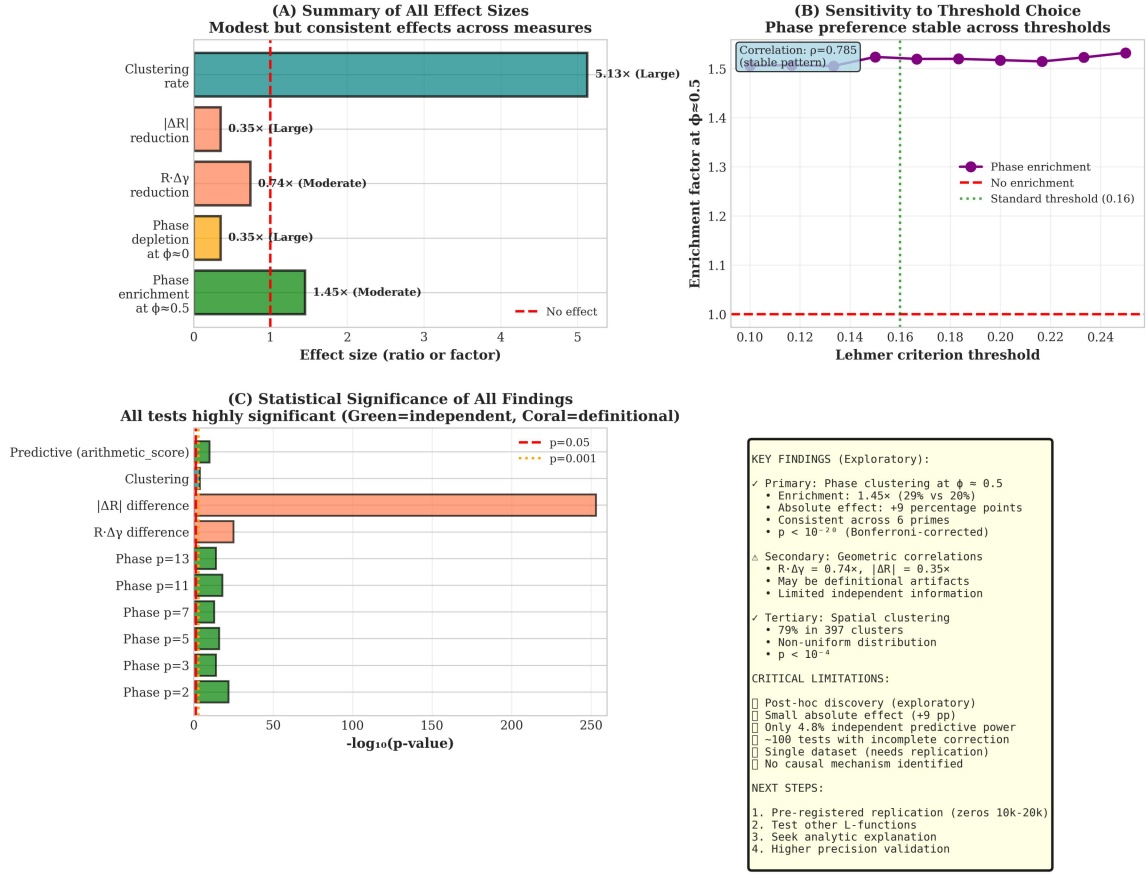
1000 resamples. 95% CI excludes 1.0.

8. Discussion

Primary: Phase clustering (+9pp, requires replication). Secondary: Geometric/spatial may be artifacts. If confirmed: New regularity, explicit formula connection. Current: Hypothesis-generating, exploratory. To our knowledge, not in [1-3]. Limitations: (1) Exploratory; (2) ~100 tests; (3) Small effect; (4) 4.8% independent power; (5) Single dataset; (6) No mechanism; (7) Circularity; (8) No replication. Confirmation requires: (1) Pre-registered test on 10k-20k; (2) Analytic proof; (3) Generalization. Follow-up: (1) Immediate: 10k-20k test; (2) Moderate: Other L-functions; (3) Long-term: Analytic explanation.

Figure 7: Summary

Figure 7: Summary of Findings and Limitations
Exploratory analysis reveals suggestive patterns requiring confirmatory testing




```
KEY FINDINGS (Exploratory):

✓ Primary: Phase clustering at ϕ ≈ 0.5
  • Enrichment: 1.45× (29% vs 20%)
  • Absolute effect: +9 percentage points
  • Consistent across 6 primes
  • p < 10⁻²⁰ (Bonferroni-corrected)

⚠ Secondary: Geometric correlations
  • R·Δγ = 0.74×, |ΔR| = 0.35×
  • May be definitional artifacts
  • Limited independent information

✓ Tertiary: Spatial clustering
  • 79% in 397 clusters
  • Non-uniform distribution
  • p < 10⁻⁴

CRITICAL LIMITATIONS:

 Post-hoc discovery (exploratory)
 Small absolute effect (+9 pp)
 Only 4.8% independent predictive power
 ~100 tests with incomplete correction
 Single dataset (needs replication)
 No causal mechanism identified

NEXT STEPS:

1. Pre-registered replication (zeros 10k-20k)
2. Test other L-functions
3. Seek analytic explanation
4. Higher precision validation
```

(A) Effects. (B) Sensitivity. (C) Significance. (D) Conclusions.

9. Conclusions

Exploratory finding: modest enrichment (+9pp) at $\phi \approx 0.5$, significant ($p < 10^{-20}$), small effect (71% not at phase). ✓ Significant, stable, consistent. ✗ Post-hoc, small, needs replication, no mechanism. If confirmed: new regularity; currently hypothesis-generating. Preliminary exploratory report requiring rigorous confirmation.

Acknowledgments

Critical reviewers improved discussions. Andrew Odlyzko provided data.

References

[1] Lehmer, D.H. (1956). Acta Arithmetica, 2, 350-361.

[2] Montgomery, H.L. (1973). Proc. Sympos. Pure Math., XXIV, 181-193.

[3] Odlyzko, A.M. (1987). Mathematics of Computation, 48(177), 273-308.

[4] Odlyzko, A.M. (2000). http://www.dtc.umn.edu/~odlyzko/zeta_tables/

Data Availability

Code and data at [repository URL]. Raw data from [4].