International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V72I10P111

Volume 72 Issue 10, 125-134, October 2026
© 2026 Seventh Sense Research Group®

Original Article

Measuring the Second-Order Constants of the Lemke Oliver–Soundararajan Bias Across Prime Moduli

Devyansh Garg

Grade 12, The Shri Ram School, Moulsari, Gurugram, Haryana, India.

Corresponding Author : gargdevyansh10@gmail.com

Received: 30 August 2026 Revised: 03 October 2026 Accepted: 19 October 2026 Published: 30 October 2026

Abstract *- Lemke Oliver and Soundararajan conjectured that the frequency of consecutive-prime residue patterns modulo q has an asymptotic expansion whose leading correction depends only on whether the two residues are equal, with all finer structure carried by a second constant given by a character sum. Earlier computations compared pattern counts with an integrated form of this conjecture. This study instead fits the second constant itself, for all 864 residue pairs of the prime moduli 3, 7, 11, 13, 17 and 19, by enumerating consecutive primes to 10^{11}, fitting disjoint increments between checkpoints so that nested counts do not understate the uncertainty. The pattern predicted by the theory is reproduced. Measured and predicted constants correlate at 0.976 for q = 7 and 0.995 for diagonal pairs, and the residue difference explains 92.8 to 99.2 percent of the variance in the measured constants at every modulus, as predicted. The values are not reproduced. The median relative error is 21 percent at q = 7 and 94 percent at q = 3, and pair-level correlation is lost from q = 13. At $x = 10^{11}$ the nominally subleading term is 1.9 to 4.2 times the leading one. The fitted constants drift toward the prediction as x grows for q up to 11 but not from q = 13, so the large-modulus discrepancy is consistent with slow convergence but not confirmed. The results separate what finite computation can test, the arithmetic pattern of the constants, from what it cannot yet test, their values.*

Keywords *- Analytic Number Theory, Consecutive Primes, Hardy-Littlewood Conjecture, Prime Distribution, Residue Classes.*

1. Introduction

In 2016, Lemke Oliver and Soundararajan [1] observed that consecutive primes are biased against repeating the same residue class modulo small integers. Among the first hundred million primes taken modulo 10, the diagonal patterns (a, a) occur far less often than the equidistribution heuristic predicts. Ash, Beltis, Gross and Sinnott [2] had earlier examined the frequencies of successive pairs of prime residues.

The deviations are much larger than Chebyshev's bias and do not follow it. The classical bias favours the classes 3 and 7 modulo 10, yet (3, 3) and (7, 7) occur less often than (1, 1) and (9, 9), which points to a distinct mechanism. Chebyshev's bias itself was characterised by Rubinstein and Sarnak [3] under the Generalized Riemann Hypothesis and a linear-independence hypothesis on the zeros, and [4] surveys the resulting theory of prime number races. Large biases also appear in other arithmetic settings, for example among products of two primes [5]. The explanation offered in [1], derived heuristically from the Hardy-Littlewood prime k-tuples conjecture [6], is that the pattern counts have the expected density in the limit but carry large secondary terms that tend to zero only slowly. The same singular-series heuristics underlie related work on the variance of primes in short intervals [7].

Qualitative questions about consecutive primes in residue classes are hard but partly settled. Shiu [8] proved that the fully diagonal pattern occurs infinitely often for any modulus and any reduced residue. Maynard [9] later showed that such strings occur with positive density. Related results on consecutive primes in tuples are given in [10], and Lau [11] has established results on which prescribed blocks of residue classes must appear among consecutive primes for squarefree modulus. None of these results speaks to how often the permissible patterns occur, which is where the bias lives. Two strands of later work bear directly on those frequencies. Lemke, Oliver and Soundararajan [12] returned to the secondary constant and related its distribution to the Fourier transform of classical Dedekind sums and to the error term in the asymptotic formula for the summatory Euler totient function. They also showed that the leading difference-only part of the constant is an odd function of the residue difference. David, Devin, Nam and Schlitt [13] constructed the analogous model for consecutive sums of two squares and recovered biases of the same shape. Separately, Holt [14] argues that the biases follow from the way gap populations evolve under the sieve of

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Eratosthenes and are ultimately transient. On that account, they are artifacts of sieve dynamics rather than the shadow of a fixed arithmetic constant.

The second-order constants themselves have not been measured systematically. The computations in [1] cover the moduli 3, 4, 5, 8, 10 and 12, most of them composite, and compare pattern counts with a numerically integrated form of the conjecture rather than estimating its constants. This paper fits the second-order constant directly for every admissible pair of the prime moduli 3, 7, 11, 13, 17 and 19, 864 pairs in all, out to $x = 10^{11}$. It differs from earlier work in four ways. It estimates the constants one pair at a time rather than comparing totals. It covers prime moduli up to 19, where the predicted substructure in the residue difference is rich enough to test. It tests that substructure with a positive control, so that a null result cannot come from an underpowered design. Moreover, it separates agreement in pattern from agreement in value, and measures how far each holds.

Three questions are asked. Do the measured constants reproduce the singular-series prediction? Do they show the substructure in the residue difference and in quadratic-residue status that the character-sum formula implies? Do any deviations discriminate between the Hardy-Littlewood account and the sieve-evolution account?

2. Materials and Methods

2.1. The Bias Function and the Expansion Fitted

For modulus q and a permissible pair (a, b) coprime to q, define the normalised bias in Equation 1. The pair count records consecutive primes p_n and p_{n+1} with p_n at most x, p_n congruent to a and p_{n+1} congruent to b modulo q. Following [1], a pattern is indexed by its least prime.

$$B(x; q, a, b) = \frac{\pi(x;q,(a,b))}{\pi(x)/\varphi(q)^2} - 1 \tag{1}$$

The Main Conjecture of [1] states the asymptotic expansion in Equation 2, where li is the logarithmic integral and the error term is of order $(\log x)^{-7/4}$.

$$\pi(x; q, \mathbf{a}) = \frac{\mathrm{li}(x)}{\varphi(q)^r}\left(1 + c_1 \frac{\log\log x}{\log x} + \frac{c_2}{\log x} + O((\log x)^{-7/4})\right) \tag{2}$$

For the bias of Equation 1, this gives the two-term form in Equation 3.

$$B(x; q, a, b) = c_1 \frac{\log\log x}{\log x} + \frac{c_2}{\log x} + O((\log x)^{-7/4}) \tag{3}$$

The leading correction is of size loglog x / log x, not 1 / log x. Its coefficient (Equation 4) takes exactly two values per modulus: one half off the diagonal, and one half minus half the totient on it. The indicator in Equation 4 is one when a and b are congruent modulo q and zero otherwise. The diagonal versus off-diagonal split is therefore fixed analytically, and every finer distinction between patterns resides in the second constant.

$$c_1(q; (a, b)) = \frac{\varphi(q)}{2}\left(\frac{1}{\varphi(q)} - 1[a \equiv b]\right) \tag{4}$$

A single-constant model, in which the bias decays purely as one over log x, forces a slowly varying loglog x into a constant. Section 3.1 quantifies the resulting misspecification.

2.2. Theoretical Baseline

For prime q and distinct residues, [1] gives the second-order constant as the character sum in Equation 5, summed over non-principal Dirichlet characters modulo q.

$$c_2(q; (a, b)) = \frac{1}{2}\log\frac{2\pi}{q} + \frac{q}{\varphi(q)}\sum_{\chi \neq \chi_0} C_{q,\chi}\left[\chi(b - a) + \frac{\chi(b) - \chi(a)}{\varphi(q)}\right] \tag{5}$$

The coefficient C(q, chi) is the product of the Dirichlet L-function at zero, the L-function at one, and an Euler product A(q, chi) over primes not dividing q (Equation 6). It vanishes unless the character is odd, since the L-function at zero vanishes for even characters.

$$C_{q,\chi} = L(0, \chi)L(1, \chi)A_{q,\chi}, \quad A_{q,\chi} = \prod_{p \nmid q}\left(1 - \frac{(1 - \chi(p))^2}{(p-1)^2}\right) \tag{6}$$

On the diagonal, the constant is independent of the residue and has a closed form. Reference [1] gives it for general modulus, and [12] gives the prime-modulus case in Equation 7.

$$c_2(q;(a,a)) = \frac{q-2}{2}\log\frac{q}{2\pi} \quad (7)$$

Two features of Equation 5 drive the analysis. First, the term in the residue difference makes dependence on that difference a prediction of the theory rather than an open question. Second, because even characters are annihilated, the leading difference-only part is an odd function of the difference, as established in [12]. The smaller term adds a dependence on the two residues individually.

Equation 5 was evaluated numerically for all 864 pairs. Dirichlet characters modulo a prime were built from a primitive root. The L-function at zero was evaluated as a finite character sum, and the L-function at one through the digamma function, because the Hurwitz zeta expression is numerically unusable at that point. The Euler product A(q, chi) was truncated at $P = 10^6$. To check this, the product was recomputed for $P = 10^4$, 10^5, 10^6, 10^7 and 10^8. Between $P = 10^6$ and $P = 10^8$ the predicted constants change by at most 1.6×10^{-6} over all 864 pairs, with a median change of 3.6×10^{-7}, and the change falls by about a factor of ten for each tenfold increase in P (Online Resource 1, Table S2). This is consistent with the bound on the neglected tail, of order 4 / (P log P), or about 3×10^{-7} at $P = 10^6$. It is also more than three orders of magnitude below the smallest standard error on a measured constant (Section 2.5). Four further checks validate the implementation: the L-function at one for the odd character modulo 3 matches its classical closed form to eighteen digits, the character sum matches Equation 7 to machine precision at every modulus, the symmetry relating a pair (a, b) to its partner (-b, -a) holds in all 864 pairs, and the difference-only part is odd to 10^{-12}.

2.3. Enumeration

Consecutive prime pairs were enumerated with a segmented sieve of Eratosthenes over NumPy boolean arrays [15]. Each segment yields a sorted array of primes. Residues are taken with one vectorised operation per modulus, and each consecutive pair is encoded as a single integer and accumulated by histogram. All six moduli are counted in a single pass. The residue of the last prime in each segment is carried forward, so pairs straddling a segment boundary are counted exactly once. This design runs approximately 380 times faster end to end than a per-prime interpreted loop over the six moduli, and needs no external prime-generation library. The original enumeration in [1] used C++ and the primesieve library [16], so the agreement in Section 2.4 is between two independent implementations using different algorithms. The full run to $x = 10^{11}$ took 685 seconds on a single core of an Apple M4 processor and gave $\pi(10^{11}) = 4{,}118{,}054{,}813$, in exact agreement with the known value.

2.4. Calibration

Counters were recorded at 31 logarithmically spaced checkpoints, six per decade, from 10^6 to 10^{11}. Correctness was checked at three levels. First, exact anchors: the prime counting function at 10^6 through 10^9 matches known values, and the vectorised counter reproduces an independent brute-force implementation pair by pair with zero discrepancies. Second, internal invariants: counts do not change with segment size, including a prime segment length chosen to place seams irregularly; cumulative counts are monotone; and per-block counts reconcile exactly with checkpoint totals. Third, external replication against [1]: across 26 published cells spanning moduli 3, 5, 8 and 12 at $x = 10^9$, the counts agree to a maximum relative error of 4.01×10^{-4} and a mean of 9.75×10^{-5}. Since [1] prints four significant figures, half a unit in the last place is 5×10^{-4}. Every cell falls inside that bound, so the discrepancy is within their rounding.

2.5. Fitting and Uncertainty

Because Equation 4 fixes the leading coefficient, subtracting the leading term leaves a quantity proportional to one over log x and linear in the single unknown second constant. Fixing that coefficient is necessary. Over the range studied, the two basis functions have sample correlation 0.999, and fitting both freely inflates the uncertainty on each by more than an order of magnitude and yields no usable constraint. Fits use the 19 checkpoints from 10^8 to 10^{11}.

The checkpoint counts are cumulative. The count at 10^{11} contains every pair counted at 10^{10}, so a fit that treats checkpoints as independent understates its uncertainty. The fit therefore uses disjoint increments. For each pair, the first observation is the count up to 10^8, and each further observation is the number of new pairs between consecutive checkpoints.

Writing the expected count up to x as $\pi(x)(1 + B(x)) / \text{phi}(q)^2$, with B given by the two-term expansion, makes each expected increment linear in the second constant. The increments cover disjoint stretches of the prime sequence, so they are treated as independent and fitted by weighted least squares with binomial variances. This is equivalent to generalised least squares on the cumulative counts with the covariance of nested counts.

Two checks set the size of the uncertainty. First, the binomial variance is conservative for sampling noise. Over 39 disjoint blocks of equal width from 2.5×10^9 to 10^{11}, the scatter of each pair's share of the counts about a smooth trend in x is 0.27 to

0.76 times the binomial value (median by modulus), because consecutive pairs share primes and their counts are constrained. Second, the two-term model does not fit the increments within binomial errors. The median chi-square per degree of freedom is 8.0 to 34.4 by modulus, which reflects the neglected higher-order terms (Section 3.3). Every reported standard error is therefore the binomial standard error multiplied by the square root of the chi-square per degree of freedom. The medians range from 0.0024 at q = 3 to 0.042 at q = 19. A block bootstrap over equal stretches of the prime sequence [17] was considered and not used. The bias changes with x, so equal stretches of x are not interchangeable, and resampling them mixes that trend into the uncertainty.

2.6. Reproducibility and Data

The enumeration, fitting and analysis code is written in Python using NumPy [15], pandas, mpmath and SymPy, and runs on a single processor core. The per-checkpoint counts for all 864 pairs (26,784 rows), the per-block counts used in Section 2.5, the fitted and predicted constants, and the code that produces every table and figure are available at https://github.com/Devonkedev/los-bias-prime-moduli.

Online Resource 1 lists, for every pair, the predicted first and second constants, the measured second constant, its standard error, the chi-square per degree of freedom and the residual (Table S1). The run reported here used a MacBook Air with an Apple M4 processor and 16 GB of memory, running macOS 26.1 and Python 3.13. The package versions that reproduce every table and figure are listed in the repository (requirements.txt).

3. Results and Discussion

3.1. The One-Term Model is Misspecified

Fitting a single-constant model separately in four consecutive windows of the fitted range gives a constant that drifts. The median spread of the fitted value across windows is 0.693, against 0.557 for the second constant from the two-term fit, and the median per-pair ratio of the two spreads is 1.11. The drift is modest because the two basis functions are nearly collinear in this range. That is why the one-term fit appears adequate while estimating a quantity that corresponds to no constant in Equation 2.

3.2. Measured Constants Against Prediction

Figure 1 shows the measured bias against one over log x for the four pairs at modulus 3, with the two-term fit and the prediction.

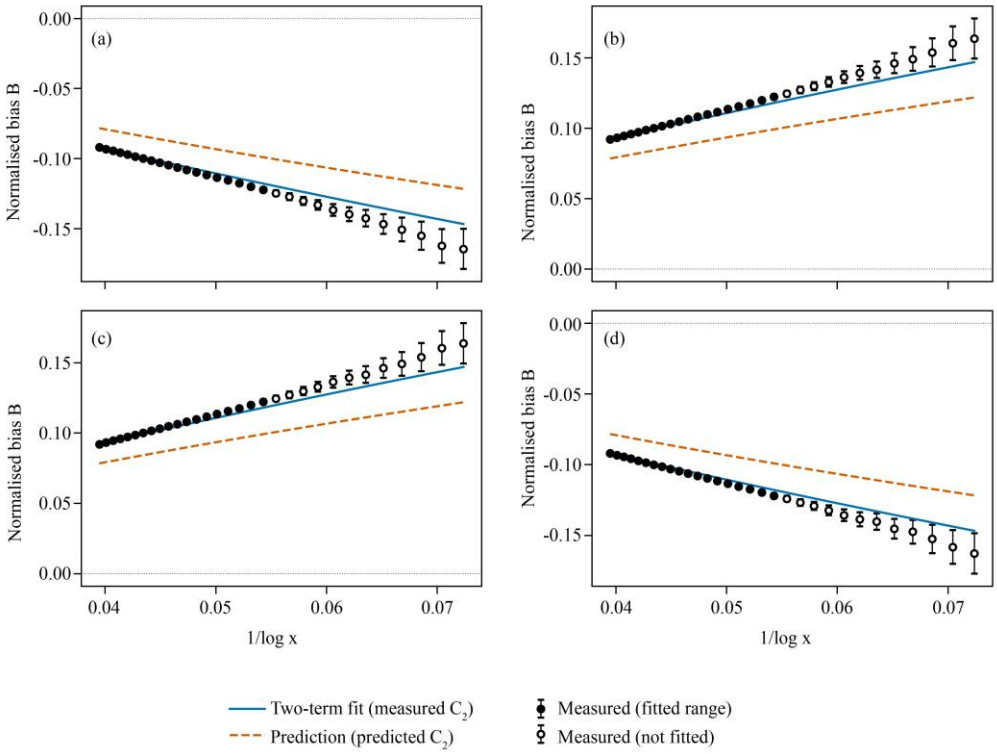

Fig. 1 Measured normalised bias against one over log x for the four residue pairs modulo 3: (a) (1, 1), (b) (1, 2), (c) (2, 1), (d) (2, 2). Filled points are in the fitted range $x \geq 10^8$. Error bars are ±2 binomial standard errors. Solid line: two-term fit with the measured second constant. Dashed line: prediction with the predicted second constant

Table 1 compares measured and predicted constants. Correlation and agreement have to be read separately. Correlation measures whether the pattern of constants across pairs is right, and the relative error measures whether the values are right. The pattern is reproduced at small modulus. At q = 7 the correlation is 0.976, and over the 64 diagonal pairs, where Equation 7 applies, it is 0.995. The values are not reproduced. At q = 7 the median relative error is 21 percent, and at q = 3 the measured constants are about 1.94 times the predicted ones. The q = 3 correlation of 1.000 carries no information, because its four pairs take only two values of opposite sign. Diagonal pairs are closest in value, with a median relative error of 7 percent. From q = 13, the correlation is consistent with zero. Figure 2 shows all 864 pairs against prediction.

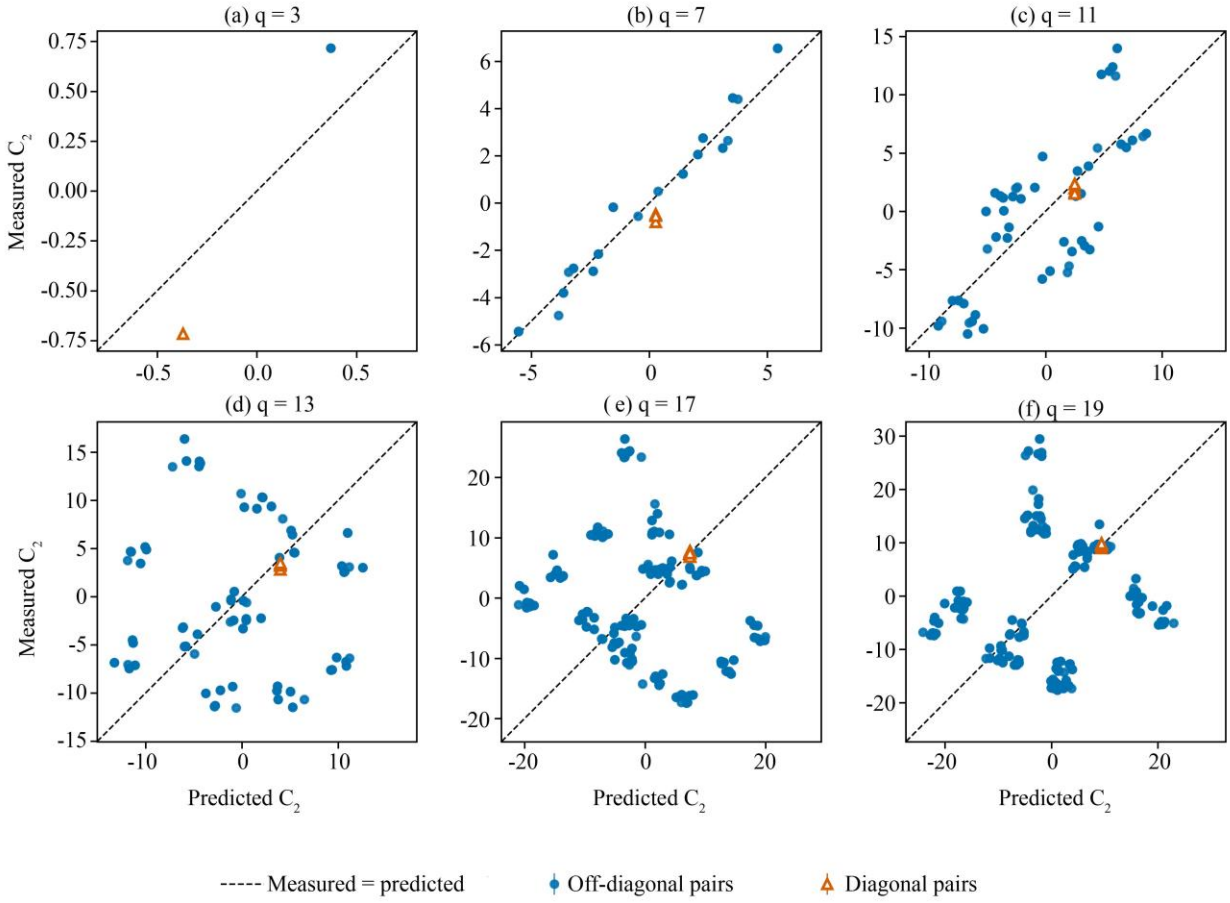

Fig. 2 Measured against predicted second-order constants for all 864 residue pairs, one panel per modulus. Error bars are ±2 reported standard errors and are smaller than the markers. Dashed line: measured equals predicted

Table 1. Measured against predicted second-order constants, by modulus

q	Pairs	Correlation	Mean |Δ|	Max |Δ|	Median relative error	Median SE
3	4	+1.000	0.346	0.346	0.94	0.0024
7	36	+0.976	0.565	1.354	0.21	0.0066
11	100	+0.749	3.311	7.861	0.57	0.0205
13	144	-0.032	7.907	22.361	1.50	0.0202
17	256	-0.212	12.156	29.749	1.32	0.0338
19	324	+0.148	12.317	31.728	0.94	0.0416
All	864	-	9.947	31.728	-	-
Diag.	64	+0.995	0.514	-	0.07	-

Δ is the measured minus the predicted constant. Relative error is |Δ| divided by the absolute predicted constant. SE is the reported standard error (Section 2.5). The largest single residual is at q = 19, (a, b) = (15, 2): measured 29.49 against predicted -2.24.

3.3. Size of the Discrepancy and the Neglected Terms

The residuals in Table 1 are not statistical noise. By modulus, their median is 67 to 368 times the reported standard error. They are also not, by themselves, evidence against Equation 2, because the two correction terms in Equation 3 have very different sizes in the accessible range.

The off-diagonal leading coefficient is one half at every modulus, while by Equation 7 the second constant grows like q log q, with off-diagonal values typically of size q. At $x = 10^{11}$, loglog x / log x is 0.128 and one over log x is 0.0395. The ratio of the second term to the first is therefore the second constant divided by the product of the first and loglog x. For the median off-diagonal pair, this ratio is 0.23 at q = 3 but 1.92, 2.63, 3.20, 3.75 and 4.16 at q = 7, 11, 13, 17 and 19. Above q = 3, the nominally subleading term is the larger of the two at the largest x reached.

The two terms cross where loglog x equals the absolute ratio of the second constant to the first, that is at $x = \exp(\exp(|c_2 / c_1|))$. For the median off-diagonal pair at q = 7, the ratio is 6.21, which gives x of about 10^{217}. For the largest q = 7 pair, it is about $10^{28,119}$, and for the median pair at q = 11, about $10^{2,116}$. This crossover is not convergence of the expansion. It marks only where the leading term overtakes the second.

The fitted second constant also absorbs the neglected term of order $(\log x)^{-7/4}$, whose coefficient is not predicted, and convergence requires that term to be small next to the second term. The crossover shows that the accessible range is far from the asymptotic regime. It does not say when the measured constants will reach their predicted values.

A direct test probes the neglected term. With both constants fixed at their predicted values and only a coefficient of $(\log x)^{-7/4}$ fitted, the maximum absolute residual over all checkpoints falls to 4×10^{-4} at q = 3 and 1.0×10^{-2} at q = 7, with coefficients of order 4 and 6. At q = 13 and above, the same procedure needs coefficients of order 65 to 138, large enough to indicate that a single next-order term no longer suffices—section 3.7 tests directly whether the discrepancy shrinks as x grows.

3.4. Substructure: Design and a Design that Fails

The natural first design regresses the off-diagonal constants on four predictors: the two lowest-frequency circular harmonics of the residue difference, and the Legendre symbols of the two residues, which encode quadratic-residue status. The circular encoding suits a difference that lives on a cycle, and the harmonics superficially resemble the difference term of Equation 5.

That design fails its own positive control, and this is reported because it would otherwise propagate into a false null. Applied to the theoretical constants, where difference dependence is present by construction, it yields a coefficient of determination of 0.14 at q = 7 and 0.006 at q = 17 and 19. Its nulls would therefore carry no information. The reason is in Equation 5. The constant is a sum over all odd characters, so its difference dependence spreads over many frequencies, and two harmonics capture an arbitrarily small share.

The primary analysis therefore enters the residue difference as a categorical factor, with one indicator per difference class. Three targets are regressed per modulus: the predicted constant as a positive control, the measured constant, and the residual between them. Diagonal pairs are excluded because the constant is fixed there. This gives fifteen omnibus tests, and probabilities are adjusted by the Benjamini-Hochberg procedure [18] to control the false discovery rate.

Three checks guard against overstating these tests. Each pair (a, b) and its partner (-b, -a) have the same predicted constant and the same residue difference, so every test is repeated with one observation per symmetric pair. Overfitting is checked by repeated five-fold cross-validation, in which each held-out pair is predicted by the mean of its class among the remaining pairs. Significance is also checked by a permutation test that shuffles the class labels 10,000 times.

3.5. Substructure Results

Table 2 gives the regressions. The residue difference explains 92.8 to 99.2 percent of the variance in the measured constants at every modulus, and these figures track the positive control almost exactly. The high values are not overfitting. Cross-validated R^2 for the measured constants is 0.88 to 0.99, and no permutation of the class labels reached the observed R^2. They are also what the theory predicts.

By [12], the off-diagonal constant divided by q equals a function of the residue difference alone plus a term of order $(\log q)^2 / \sqrt{q}$, so most of its variance must lie between the difference classes, and the control rows confirm this. Counting each symmetric pair once leaves R^2 essentially unchanged and gives the adjusted probabilities in the last column, which remain far below any conventional threshold.

Table 2. Variance in the second-order constant explained by residue-difference class

q	Target	Pairs	R^2	CV R^2	Perm. p	Sym. pairs	Adj. p (sym.)
7	Predicted (control)	30	0.933	0.892	<1e-4	18	2.0e-6
7	Measured	30	0.928	0.880	<1e-4	18	3.0e-6
7	Residual	30	0.771	0.616	<1e-4	18	1.8e-3
11	Predicted (control)	90	0.978	0.971	<1e-4	50	1.7e-29
11	Measured	90	0.976	0.968	<1e-4	50	6.8e-29
11	Residual	90	0.973	0.964	<1e-4	50	6.0e-28
13	Predicted (control)	132	0.985	0.981	<1e-4	72	3.3e-49
13	Measured	132	0.983	0.979	<1e-4	72	3.2e-48
13	Residual	132	0.984	0.980	<1e-4	72	1.3e-48
17	Predicted (control)	240	0.992	0.990	<1e-4	128	3.2e-107
17	Measured	240	0.990	0.989	<1e-4	128	4.6e-104
17	Residual	240	0.991	0.990	<1e-4	128	8.3e-106
19	Predicted (control)	306	0.993	0.993	<1e-4	162	1.4e-145
19	Measured	306	0.992	0.991	<1e-4	162	7.4e-141
19	Residual	306	0.993	0.992	<1e-4	162	1.0e-142

Control rows verify that the design detects the structure the theory contains. CV R^2: mean out-of-sample R^2 over 200 repetitions of five-fold cross-validation. Perm. p: from 10,000 shuffles of the class labels, smallest attainable value 1e-4. Last column: F-test with one observation per symmetric pair (a, b), (-b, -a), Benjamini-Hochberg adjusted across the 15 tests.

The residual between measurement and prediction is also organised by difference class, with R^2 from 0.77 at q = 7 to 0.99 at q = 17 and 19. The deviation from prediction is therefore systematic rather than scattered across pairs. One possible reading, consistent with Section 3.3, is that the neglected higher-order term carries its own dependence on the residue difference. This is a hypothesis. The present data show only that the residual is difference-structured, not what produces that structure.

No quadratic-residue effect is detectable independently of the difference. This is consistent with Equation 5, where quadratic-residue status enters only through the real odd character's contribution to the difference term, so the difference classes absorb it.

3.6. Magnitude is Recovered Even Where Assignment is Not

The pair-level correlations of Table 1 understate how much of the predicted structure is present. Table 3 groups pairs into difference classes and compares measured and predicted class means. The correlation between class-mean profiles behaves like the pair-level correlation, high at q = 7 (0.987) and lost by q = 13. The spread of the class means, which is the amplitude of the substructure, is close to the predicted spread at every modulus, with ratios from 1.01 to 1.27. The predicted class means are almost exactly antisymmetric in the difference, as [12] requires. The measured means are antisymmetric only approximately. Their antisymmetry defect is 2.9 to 4.6 times the predicted one, though still only 4 to 21 percent of the spread.

Table 3. Difference-class means: assignment versus amplitude

q	Classes	Correlation	Measured spread	Predicted spread	Ratio	Antisym. defect (measured)	Antisym. defect (predicted)
7	6	+0.987	3.605	3.412	1.06	0.157	0.054
11	10	+0.748	6.659	5.224	1.27	1.002	0.280
13	12	-0.056	7.912	7.326	1.08	1.175	0.364
17	16	-0.259	10.686	9.922	1.08	2.282	0.498
19	18	+0.113	12.025	11.893	1.01	2.017	0.553

Spread is the standard deviation of the class means. The antisymmetry defect is the mean of $|m(d) + m(q - d)| / 2$ over classes d, which is zero for an exactly odd profile.

So at q = 13 and above the measurement recovers substructure of about the right size and approximately the right antisymmetry, but cannot say which difference class receives which value. This is the pattern expected if a correctly sized predicted effect is overlaid by an unmodelled term of comparable size, the situation described in Section 3.3.

3.7. Does the Discrepancy Shrink as x Grows?

If the large-modulus discrepancy comes from slow convergence, the fitted second constant should move toward the predicted value as the fit moves to larger x. To test this, the constant was refitted separately in four consecutive windows of x, each using only its own increments, and the first window $[10^8, 4.6 \times 10^8]$ was compared with the last $[3.2 \times 10^{10}, 10^{11}]$ (Table 4).

Table 4. Movement of the fitted second constant between the first and last windows of x

q	Pairs	Moved toward prediction	Share	Median share of gap closed	Sign-test p
3	4	4	1.00	+0.13	0.062
7	36	25	0.69	+0.16	0.014
11	100	67	0.67	+0.07	4.4e-4
13	144	66	0.46	-0.01	0.861
17	256	113	0.44	-0.03	0.974
19	324	157	0.48	-0.00	0.729

A pair moved toward the prediction if its constant changed in the direction of the predicted value. Share of gap closed is the change divided by the initial difference from the prediction. Sign-test p is one-sided against a share of one half.

At q = 3, 7 and 11 the constant moves toward the prediction, in 4 of 4, 25 of 36 and 67 of 100 pairs, closing a median of 13, 16 and 7 percent of the gap. At q = 13, 17 and 19, it does not. The shares moving toward the prediction are 46, 44 and 48 percent, which is what chance would give. Slow convergence is therefore supported where the expansion is closest to its asymptotic regime, and untested where the discrepancy is largest. At q of 13 and above, it remains a hypothesis. This agrees with an observation of [1], who found for modulus 5 that the predicted ordering of two patterns is reversed for all x below 5.076×10^{11}, beyond the range studied here, and speculated that further lower-order terms are responsible.

3.8. Which Theoretical Account is Supported

The two accounts are not formally contradictory, since the secondary terms of [1] also tend to zero. They differ in what generates the constants, and so in what the data should look like. The Hardy-Littlewood account makes a point prediction for each of the 864 pairs. The sieve-evolution account of [14] makes none, and offers instead a qualitative mechanism by which biases arise transiently and fade.

The data support the Hardy-Littlewood account in pattern but not, so far, in value. The predicted pattern across pairs is reproduced at q = 7 and on the diagonals (Section 3.2), and the predicted substructure in the residue difference appears at every modulus with the predicted amplitude (Sections 3.5 and 3.6). The predicted values are not reproduced at any modulus in this range. For q of 11 or less, the measured constants move toward them as x grows (Section 3.7). At q of 13 and above, they do not, so agreement there is neither shown nor ruled out. The data are uninformative about the sieve-evolution account.

A further consideration weighs mildly in favour of the singular-series account, though it is not evidence from the present data. The Hardy-Littlewood heuristic carries over to at least one other sequence: David, Devin, Nam and Schlitt [13] construct the analogue for consecutive sums of two squares and recover biases of the same form. A mechanism tied specifically to the evolution of the sieve of Eratosthenes over the primes has no equally natural counterpart there, whereas a singular-series argument does.

The design is asymmetric. Because [14] supplies no quantitative prediction at these moduli, the results can confirm or fail to confirm the Hardy-Littlewood account but cannot positively confirm the sieve-evolution account. The reversed ordering documented by [1] (Section 3.7) gives the emphasis of [14] on transient, range-dependent effects some indirect support from within the Hardy-Littlewood programme itself. For moduli 13 and above, the constants of Equation 5 are effectively beyond direct confirmation in this range.

3.9. Comparison with Earlier Computations

The closest earlier computation is that of [1], who enumerated consecutive primes for the moduli 3, 4, 5, 8, 10 and 12 and compared pattern counts with a numerically integrated form of their conjecture. That comparison tests whether the conjecture as a whole reproduces the counts. The present study tests a different and stricter thing: the individual constants of the asymptotic form, pair by pair. Three choices make this possible. Fixing the leading coefficient at its exact value removes the near-collinearity that otherwise leaves the second constant unidentified (Section 2.5). Fitting disjoint increments gives uncertainties that respect

the nesting of the counts. Moreover, prime moduli up to 19 provide enough residue classes to test the predicted difference structure against a positive control.

The vectorised enumeration reaches 10^{11} in about eleven minutes on one core. This is an order of magnitude below the largest computation of [1], so the gain is in what is measured rather than in range. Where the two studies overlap, they agree: the counts of [1] for the moduli 3, 5, 8 and 12 are reproduced to within their printed precision (Section 2.4). The large gap found here between measured and predicted constants is consistent with the statement in [1] that the integrated form fits the counts better than the asymptotic constants. The other studies cited address the constants theoretically [12] or in a different sequence [13], so they do not give fitted values that can be compared directly.

3.10. Limitations

Five limitations bound these conclusions. First, the run reaches $x = 10^{11}$, an order of magnitude below the largest computation of [1]. Given Section 3.3, extending the range by a few orders of magnitude would not reach the asymptotic regime. Second, the reported standard errors are scaled by the misfit of the two-term model, so they describe how well that model fits as well as sampling noise. They assume that disjoint stretches of the prime sequence are independent, which is an approximation. Third, the comparison is against the asymptotic constants, whereas [1] compares counts with a numerically integrated form that they state fits better. Part of the offset at large modulus may reflect that choice. Fourth, the slow-convergence explanation is supported only for q of 11 or less (Section 3.7), and at larger moduli it is a hypothesis. Fifth, the initial two-harmonic design was underpowered (Section 3.4). The positive control exposed it, and any study of this kind should include one.

4. Conclusion

The second-order constants of the Lemke Oliver-Soundararajan conjecture were measured directly for six prime moduli and all 864 admissible residue pairs, out to $x = 10^{11}$, with uncertainties that account for the nesting of cumulative counts. The results separate qualitative from quantitative agreement. Qualitatively, the predictions are borne out. The pattern of constants across pairs is reproduced at $q = 7$ and on the diagonal pairs, and the predicted dependence on the residue difference appears at every modulus with the predicted amplitude, explaining 92.8 to 99.2 percent of the variance. Quantitatively, they are not yet confirmed. The measured constants differ from the predicted ones at every modulus, by a median of 21 percent at $q = 7$ and by more than the constants themselves at $q = 13$ and 17. For q of 11 or less, the measured constants move toward the predictions as x grows, which supports slow convergence of the expansion. For q of 13 and above, no such movement is seen in the accessible range, so slow convergence there is a hypothesis rather than a finding. The single-constant framing of the bias is also misspecified, because the leading term in loglog x over log x has a coefficient fixed by whether the two residues agree. Further work should fit the integrated form of the conjecture rather than its asymptotic constants, which would test the predicted values at the x that computation can reach.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The per-checkpoint and per-block pair counts, the fitted and predicted constants for all 864 pairs (Online Resource 1) and the code that produces every table and figure are available at https://github.com/Devonkedev/los-bias-prime-moduli.

Author Contribution Statement

DG formulated the research question, carried out the analysis, interpreted the results and wrote the manuscript. DG read and approved the final manuscript.

Acknowledgments

The author thanks Dr. Kunal Mittal for his guidance throughout this project and for substantial assistance with the computational work, and his school and his parents for their support.

References

[1] Robert J. Lemke Oliver, and Kannan Soundararajan, "Unexpected Biases in the Distribution of Consecutive Primes," *Proceedings of the National Academy of Sciences*, vol. 113, no. 31, pp. E4446-E4454, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] Avner Ash et al., "Frequencies of Successive Pairs of Prime Residues," *Experimental Mathematics*, vol. 20, no. 4, pp. 400-411, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[3] Michael Rubinstein, and Peter Sarnak, "Chebyshev's Bias," *Experimental Mathematics*, vol. 3, no. 3, pp. 173-197, 1994. [CrossRef] [Google Scholar] [Publisher Link]

[4] Andrew Granville, and Greg Martin, "Prime Number Races," *The American Mathematical Monthly*, vol. 113, no. 1, pp. 1-33, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[5] David Dummit, Andrew Granville, and Hershy Kisilevsky, "Big Biases amongst Products of Two Primes," *Mathematika*, vol. 62, no. 2, pp. 502-507, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[6] Godfrey H. Hardy, and John E. Littlewood, "Some Problems of 'Partitio Numerorum'; III: on the Expression of a Number as a Sum of Primes," *Acta Mathematica*, vol. 44, no. 1, pp. 1-70, 1923. [CrossRef] [Google Scholar] [Publisher Link]

[7] Hugh L. Montgomery, and Kannan Soundararajan, "Primes in Short Intervals," *Communications in Mathematical Physics*, vol. 252, no. 1-3, pp. 589-617, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[8] Daniel K.L. Shiu, "Strings of Congruent Primes," *Journal of the London Mathematical Society*, vol. 61, no. 2, pp. 359-373, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[9] James Maynard, "Dense Clusters of Primes in Subsets," *Compositio Mathematica*, vol. 152, no. 7, pp. 1517-1554, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[10] William D. Banks, Tristan Freiberg, and Caroline L. Turnage-Butterbaugh, "Consecutive Primes in Tuples," *Acta Arithmetica*, vol. 167, no. 3, pp. 261-266, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[11] Cheuk Fung Lau, "Residue Class Patterns of Consecutive Primes," *arXiv Preprint*, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Robert J. Lemke Oliver, and Kannan Soundararajan, "The Distribution of Consecutive Prime Biases and Sums of Sawtooth Random Variables," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 168, no. 1, pp. 149-169, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Chantal David et al., "Lemke Oliver and Soundararajan Bias for Consecutive Sums of Two Squares," *Mathematische Annalen*, vol. 384, no. 3-4, pp. 1181-1242, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Fred B. Holt, "Expected Biases in the Distribution of Consecutive Primes," *arXiv Preprint*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Charles R. Harris et al., "Array Programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357-362, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Kim Walisch, Primesieve: Fast Prime Number Generator, Software Library, 2024. [Online]. Available: https://github.com/kimwalisch/primesieve

[17] Bradley Efron, and Robert J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall/CRC, New York, 1st ed., pp. 1-436, 1993. [CrossRef] [Google Scholar] [Publisher Link]

[18] Yoav Benjamini, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289-300, 1995. [CrossRef] [Google Scholar] [Publisher Link]