RANK BIAS FOR ELLIPTIC CURVES MOD p

KIMBALL MARTIN AND THOMAS PHARIS

ABSTRACT. We conjecture that, for a fixed prime p, rational elliptic curves with higher rank tend to have more points mod p. We show that there is an analogous bias for modular forms with respect to root numbers, and conjecture that the order of the rank bias for elliptic curves is greater than that of the root number bias for modular forms.

1. INTRODUCTION AND CONJECTURES

An elliptic curve E over a field k is a smooth projective curve of genus 1 with a distinguished k-point O. If the characteristic of k is not 2, then E has Weierstrass model $y^2 = f(x)$, where $f(x)$ is a cubic polynomial with distinct roots, and O is the point at infinity. Moreover, E has the structure of an abelian group, with the distinguished point O being the identity element. Mordell showed that a rational elliptic curve $E/\mathbb{Q}$ is a finitely-generated group, and thus the set of rational points $E(\mathbb{Q})$ is isomorphic as an abstract group to $\mathbb{Z}^r \oplus E_{\text{tors}}(\mathbb{Q})$ for some $r \geq 0$, where $E_{\text{tors}}(\mathbb{Q})$ is the finite abelian group of rational torsion points. The non-negative integer r is called the (algebraic) rank of E.

A natural question is whether elliptic curves with more rational points have more points mod p. To be precise, any rational elliptic curve $E/\mathbb{Q}$ has a minimal model $y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$ over $\mathbb{Z}$. Minimality means that the discriminant $\Delta \in \mathbb{Z} - \{0\}$ of this equation is minimal in absolute value. For a prime $p \nmid \Delta$, one can reduce E mod p to get an elliptic curve $E(\mathbb{F}_p)$ over $\mathbb{F}_p$. The expected number of points on an elliptic curve over $\mathbb{F}_p$ is $p+1$, and Hasse's bound states that the deviation from the mean, $a_p = a_p(E) = p + 1 - \#E(\mathbb{F}_p)$, is less than $2\sqrt{p}$ in absolute value. Our question then is: for $p \nmid \Delta$, does a_p tend to be smaller, i.e., does $\#E(\mathbb{F}_p)$ tend to be larger, when E has larger rank?

Much is understood in the "horizontal" direction. Namely if E is fixed, then the Sato–Tate conjecture (now a theorem by [BLGHT11]) asserts that the limiting distribution of $\frac{a_p}{\sqrt{p}}$ is independent of the rank of E. However, if one numerically computes examples for various E, one notices an apparent bias—the a_p's tend to be smaller when E has larger rank. Necessarily such a bias would be bounded by the error term in the convergence to the Sato–Tate distribution. Indeed, one may interpret the Birch and Swinnerton-Dyer conjecture as a certain measure of this bias. In its original formulation in [BSD65], it asserts that, up to a constant, the order of growth of $\prod_{p<X} \frac{\#E(\mathbb{F}_p)}{p}$ is $(\log X)^r$, where p runs over primes of good reduction (i.e., $p \nmid \Delta$) and r is the rank of E. This may be very loosely interpreted as saying that elliptic curves with higher ranks

have more points mod p for large primes p. We also remark that Nagao's conjecture (see [Nag97], [RS98]) makes a similar prediction for averages over 1-parameter families of elliptic curves.

Here we investigate the above question in the "vertical" direction: for a fixed prime p and varying elliptic curves E of rank r, does a_p tend to be smaller the larger r is? Or to put it loosely: do elliptic curves with higher ranks also have more points mod p for small primes p? A basic issue is how to try to measure such a bias, as the Birch and Swinnerton-Dyer framework (essentially a weighted geometric mean of a_p's as p varies) has no obvious analogue for a fixed p.

Since such problems seem very difficult to tackle theoretically, we investigate this question computationally, and theoretically study an analogous question for modular forms. We will first discuss the case of modular forms, which will help motivate our framework for measuring and conjecturing bias for elliptic curves.

We remark that, following the vein of Nagao's conjecture, different questions about biases of a_p's in the vertical direction have been studied in 1-parameter families in a number of recent articles (e.g., [MMRW16], [ACF+], [KN]).

1.1. **Root number bias for modular forms.** Elliptic curves E of conductor N correspond to rational newforms $f \in S_2(N)$ of weight 2 and level N such that $L(s, E) = L(s, f)$, and thus each $a_p(f) = a_p(E)$. The root number w of f (or E) is ± 1, which is the sign in the functional equation of the L-function. The analytic rank of this L-function is even or odd according to whether w is $+1$ or -1. According to the Birch and Swinnerton-Dyer conjecture, the analytic rank should be the same as the algebraic rank. The minimalist conjecture predicts that 100% of the time, the rank is 0 or 1, according to the root number being $+1$ or -1.

In Corollary 2.3, we show that for a fixed prime p, the average of $a_p(f)$ over newforms $f \in S_2(N)$ with root number ± 1 grows approximately like $\pm\frac{1}{\sqrt{N}}$ for large squarefree N prime to p. In fact, we more generally treat newforms of even weight k, and the method applies to arbitrary Fourier coefficients $a_n(f)$, however the signs for $k \equiv 0 \bmod 4$ are opposite to those for $k \equiv 2 \bmod 4$.

We now recast this bias in terms of weighted averages of Fourier coefficients over varying N. Let $\mathcal{F}^\pm(X)$ be the union of the sets of newforms f with root number ± 1 in $S_2(N)$, as N ranges over squarefree levels less than X. For a newform f, let N_f denote its exact level, i.e., f is a newform in $S_2(N_f)$. Let $\phi : \mathbb{N} \to \mathbb{R}_{>0}$ be a monotonic non-decreasing function of at most polynomial growth. We call such a ϕ a *weight function*. Consider the weighted average of p-th Fourier coefficients,

$$(1.1) \qquad \mathcal{A}^\pm(p, X; \phi) = \frac{1}{\#\mathcal{F}^\pm(X)} \sum_{f \in \mathcal{F}^\pm(X)} a_p(f)\phi(N_f).$$

Then Corollary 2.3 implies the following.

Proposition 1.1. *Fix a prime p, a weight function ϕ as above, and $\epsilon > 0$.*
  (1) For any ϕ, $\mathcal{A}^+(p, X; \phi) > \mathcal{A}^-(p, X; \phi)$ for sufficiently large X.
  (2) If $\phi(N) \ll N^{\frac{1}{2}-\epsilon}$, then $\mathcal{A}^\pm(p, X; \phi) \to 0$ as $X \to \infty$.
  (3) If $\phi(N) \gg N^{\frac{1}{2}+\epsilon}$, then $\mathcal{A}^\pm(p, X; \phi) \to \pm\infty$ as $X \to \infty$.

The polynomial growth condition on ϕ is not actually needed for this proposition, but we impose it for the purposes of preventing erratic behavior in weighted averages for sequences with greater variation. The reason to restrict to forms of squarefree level is to avoid alternating sums of class numbers that arise in the trace formula for non-squarefree levels.

The first statement asserts that there is a persistent root number bias in the a_p's. (In fact, for any ϕ, $\mathcal{A}^+(p, X; \phi) > 0$ and $\mathcal{A}^-(p, X; \phi) < 0$ for sufficiently large X, which corresponds to the sign matching in assertion (3).) The latter two statements quantify the size of the bias—they say the bias is roughly on the order of the inverse square root of the level. While Proposition 1.1 is less precise than Corollary 2.3, this formulation provides a model for investigating and measuring bias for elliptic curves.

Our original question about the rank elliptic curves affecting the size of the a_p's is geometrically motivated. One can also interpret the bias in Proposition 1.1 geometrically as follows.

Let $f \in S_2(N)$ be a newform with rationality field K_f. Its rationality degree $d = [K_f : \mathbb{Q}]$ equals the number of Galois conjugates f^σ of f. Now (the Galois orbit of) f corresponds to a rational abelian variety A_f of dimension d such that $L(s, A_f) = \prod_\sigma L(s, f^\sigma)$. Let $a_p(A_f) = \sum_\sigma a_p(f^\sigma) = \operatorname{tr}_{K_f/\mathbb{Q}} a_p(f)$. Then, for $p \nmid N$, $\#A_f(\mathbb{F}_p) = \prod_\sigma (p + 1 - a_p(f^\sigma))$, which is a degree d monic polynomial in p, and the coefficient of p^{d-1} is $-a_p(A_f)$. Hence we may think of $a_p(A_f)p^{d-1}$ being a "first order estimate" for $p^d - \#A_f(\mathbb{F}_p)$.

For simplicity, say N is prime. Then $S_2(N)$ has 2 Atkin–Lehner eigenspaces, each of size approximately $\frac{N-1}{24}$ (e.g., see [Mar18]), corresponding to root numbers $+1$ and -1. We conjectured in [Mar21] that each Atkin–Lehner space is generated by a single Galois orbit 100% of the time. Suppose the root number $+1$ (resp. -1) newforms are all Galois conjugates of a single newform f_+ (resp. f_-). The new part of the Jacobian $J_0(N)$ of $X_0(N)$ decomposes into 2 simple pieces $A_+ \oplus A_-$, where $A_\pm = A_{f_\pm}$, and the average of the $a_p(f)$'s with root number ± 1 is approximately $\frac{24}{N-1} \cdot a_p(A_\pm)$. By the minimalist philosophy, we further expect that 100% of the time A_+ has rank 0 and A_- has rank $\dim A_- \approx \frac{N-1}{24}$. Now Proposition 1.1 (or rather Corollary 2.3 to restrict to prime levels) says that, for fixed $p \nmid N$ and prime $N \to \infty$, $a_p(A_\pm)$ grows roughly like $\pm\sqrt{N}$. In particular, the $a_p(A)$'s tend to be smaller for higher-rank modular (GL(2) type) abelian varieties along this family. These remarks extend to the case of squarefree N with suitable modification.

1.2. Rank bias for elliptic curves.
Two common ways of counting rational elliptic curves E are to (partially) order by conductor or (a suitable choice of) height. In addition, one can either count isomorphism classes or isogeny classes of curves, but it seems likely that this distinction will not significantly affect the statistics we consider. Note that $a_p(E)$ only depends on the isogeny class of E.

For definiteness, we let $\mathcal{E}$ be one of the following three families of (partially ordered classes of) rational elliptic curves: (i) isogeny classes of elliptic curves of prime conductor, (partially) ordered by conductor; (ii) isogeny classes of all elliptic curves, ordered by conductor; and (iii) isomorphism classes of all elliptic curves, ordered by (minimal) naive height. These families are respectively denoted by $\mathcal{E}_{\mathrm{pr}}$, $\mathcal{E}_{\mathrm{all}}$, and $\mathcal{E}_{\mathrm{ht}}$.

When $\mathcal{E} = \mathcal{E}_{\mathrm{pr}}$ or $\mathcal{E}_{\mathrm{all}}$, let $|E| = N_E$ denote the conductor of E. When $\mathcal{E} = \mathcal{E}_{\mathrm{ht}}$, let $|E| = H_E$ denote the minimum naive height in the isomorphism class of E. Let $\mathcal{E}(X)$ be the set of (classes of) elliptic curves in $\mathcal{E}$ with $|E| < X$. Let $\mathcal{E}_r(X)$ be the subset of E in $\mathcal{E}(X)$ of rank r. Set $\mathcal{E}_r = \bigcup_X \mathcal{E}_r(X)$.

For a weight function ϕ as above, we define the weighted average

$$(1.2) \qquad \mathcal{A}_r(p, X; \phi) = \mathcal{A}_r^{\mathcal{E}}(p, X; \phi) = \frac{1}{\#\mathcal{E}_r(X)} \sum_{E \in \mathcal{E}_r(X), |E| < X, p \nmid N_E} a_p(E)\phi(N_E).$$

(The condition $p \nmid N_E$ is equivalent to $p \nmid \Delta$, where Δ is the minimal discriminant.) Note that in all cases we are weighting by conductor, in analogy with the weighting of modular forms by level in (1.1). When ordering by height H_E, one could instead weight the averages by $\phi(H_E)$ rather than $\phi(N_E)$, and we briefly remark on this below.

The following two conjectures assert an analogue of Proposition 1.1 for quantifying a rank bias in the a_p's of elliptic curves for any fixed prime p.

Conjecture 1.2 (Existence of rank bias). *Let $\mathcal{E}$ be $\mathcal{E}_{\mathrm{pr}}$, $\mathcal{E}_{\mathrm{all}}$ or $\mathcal{E}_{\mathrm{ht}}$, and $r_1 \leq r_2$. Suppose $\mathcal{E}_{r_1}$ and $\mathcal{E}_{r_2}$ are infinite. Then for any weight function ϕ, $\mathcal{A}_{r_1}(p, X; \phi) > \mathcal{A}_{r_2}(p, X; \phi)$ for all sufficiently large X.*

Conjecture 1.3 (Order of rank bias). *Let $\mathcal{E}$ be $\mathcal{E}_{\mathrm{pr}}$, $\mathcal{E}_{\mathrm{all}}$ or $\mathcal{E}_{\mathrm{ht}}$, and fix $r \geq 0$. Suppose $\mathcal{E}_r$ is infinite. Then there exists a $\delta > 0$ such that for any $\epsilon > 0$ the following hold.*

(1) If $\phi(N) \ll (\log N)^{\delta - \epsilon}$, then $\mathcal{A}_r(p, X; \phi) \to 0$ as $X \to \infty$
(2) If $\phi(N) \gg (\log N)^{\delta + \epsilon}$, then $|\mathcal{A}_r(p, X; \phi)| \to \infty$. Moreover $\mathcal{A}_r(p, X; \phi) \to +\infty$ if $r \leq 1$ and $\mathcal{A}_r(p, X; \phi) \to -\infty$ if $r \geq 2$.

The first conjecture asserts that there is a persistent bias toward the a_p's being smaller for larger ranks, when we consider sufficiently large families of elliptic curves. Note that in special families, e.g., quadratic twists, a_p's can be have in special ways which do not exhibit such a bias. It is not known for which ranks r there are infinitely many (or even any) curves of rank r (see [PPVW19] for recent conjectures). Hence we impose hypotheses on the infinitude of $\mathcal{E}_{r_1}$ and $\mathcal{E}_{r_2}$ to try to avoid considering families which are too small or special.

We remark that while it seems difficult to unconditionally prove there are infinitely many elliptic curves of rank r for $r \geq 2$ (see [BJ16] for a proof for $r = 2$ conditional on the parity conjecture), there are a number of results showing that there are infinitely many elliptic curves of rank $\geq r$, with the record being for $r = 19$ in unpublished work of Elkies (see, e.g., [PPVW19] for a summary of results in this vein).

The second conjecture asserts that the rank bias has roughly inverse polylogarithmic order in the conductor. In fact, it is plausible that the order of bias is simply inverse logarithmic, i.e., one can take $\delta = 1$ in Conjecture 1.3, but the data are not entirely clear (see Section 3). Note that if we take the constant weight function $\phi(N) = 1$, Conjecture 1.3(1) asserts that the honest averages of the a_p's tend to 0, similar to the case of modular forms in Proposition 1.1. However, there are a couple of obvious differences from the situation of Proposition 1.1.

One evident difference is that the approximate order of this bias $((\log N)^{-\delta})$ is larger than what we saw for modular forms $(N^{-\frac{1}{2}})$. From the geometric interpretation of Proposition 1.1, there is no obvious guess for the order of bias of $a_p(A)$'s for modular

abelian varieties as a simultaneous function of rank, dimension and conductor, since the dimension, conductor and rank are all strongly correlated that context. However, the notion that the order of rank bias is larger than any $N^{-\epsilon}$ when we restrict to A of bounded dimension, e.g., for elliptic curves, at least seems compatible with Proposition 1.1.

Another difference is that we are measuring rank bias for elliptic curves, but root number bias for modular forms. By the minimalist conjecture, it is natural to expect that if we look at averages of $a_p(E)$'s where E has root number $+1$ (resp. -1), this should behave the same as the rank 0 (resp. rank 1) averages. However, Conjecture 1.3(2) says that the $a_p(E)$'s tend to be positive for both rank 0 and rank 1, rather than matching the sign of the root number!

It could happen that, say, the contribution from the rank 3 and rank 5 elliptic curves is significant enough to make the (suitably weighted averages of) $a_p(E)$'s tend to be negative for root number -1. On the other hand, it also seems theoretically plausible that for either root number $+1$ or -1 (and thus for all ranks combined), the $a_p(E)$'s tend to be positive, and that the geometric reason for the root number bias in signs for modular forms is due to large rank abelian varieties. Exploratory calculations did not exhibit a positive bias in (weighted averages of) $a_p(E)$'s with fixed root number. However, it is hard to draw clear conclusions from these calculations because the numerical convergence of ranks to the minimalist conjecture is very slow, i.e., in the range we are able to compute, there is still a very large proportion of rank ≥ 2 curves (see below).

We do remark that if we count integral Weierstrass equations of elliptic curves ordered by height (which is almost the same as our family $\mathcal{E}_{\mathrm{ht}}$), then the reductions mod p are evenly distributed. By a result of Birch [Bir68], this means the unweighted averages of $a_p(E)$'s over all ranks for this family tends to 0. This is compatible with Conjecture 1.3(1), and it does not seem to preclude the possibility that suitably weighted averages of $a_p(E)$'s over all ranks may be positive.

1.3. Evidence and meta-analysis.
Our evidence for these conjectures is purely computational, and is presented in Section 3.

For $\mathcal{E}_{\mathrm{pr}}$ and $\mathcal{E}_{\mathrm{all}}$, we estimate weighted averages using the Stein–Watkins databases [SW02] consisting of over 11 milion isogeny classes of prime conductor $N < 10^{10}$ and over 115 million isogeny classes of arbitrary conductors $N \leq 10^8$. The Stein–Watkins databases do not catalogue all isogeny classes in these conductor ranges, but at least the Stein–Watkins prime conductor database appears to be nearly complete: [BGR19] estimates it contains over 99.8% of curves with prime conductor $N < 10^{10}$. (In fact [BGR19] computed a much larger database of prime conductor elliptic curves, but that database does not include rank calculations which we require.) For $\mathcal{E}_{\mathrm{ht}}$, we compute weighted averages using the height database from [BHK$^+$16], which contains all of the over 238 million curves with naive height $H \leq 2.7 \cdot 10^{10}$.

In fact, the reason for formulating our conjectures for the three specific families $\mathcal{E}_{\mathrm{pr}}$, $\mathcal{E}_{\mathrm{all}}$ and $\mathcal{E}_{\mathrm{ht}}$ is that they correspond to these existing extensive databases of curves that include conjectural ranks. The Stein–Watkins databases include numerically computed analytic ranks. The ranks computed in the height database in general assume several standard conjectures, but are unconditional over 80% of the time.

For these 3 databases, we compute weighted averages of $a_p(E)$'s with $|E| < X$ and fixed rank $r \leq 5$ for both a variety of weight functions ϕ, and a variety of primes $p \leq 300$. Note that there are not enough curves in these databases to get meaningful statistics for $r \geq 6$. For a fixed weight function ϕ and rank r, we found the general behavior to be more-or-less similar for each of the 3 databases and for any choice of p. However we will point out a couple of apparent exceptions to this trend in Section 3.

When $\phi(N) = 1$, the weighted average graphs quite quickly tend to zero as asserted in Conjecture 1.3(1), and do not appear to cross each other for different ranks beyond very small X, as asserted in Conjecture 1.2. When $\phi(N) \gg \sqrt{N}$ or larger, the weighted averages also clearly tend to $\pm\infty$. More generally, for $\phi(N) = N^\delta$, the weighted average graphs appear to grow like $CX^\delta(1 + O(X^\epsilon))$ for some constant nonzero C. For $\phi(N) = \log(N)$, some graphs appear as though they may have a finite nonzero asymptote, and some as though they may be slowly increasing or decreasing, depending on both p and r. For $\phi(N) = \log \log N$ most graphs appear to go to 0, and for $\phi(N) = (\log N)^2$ most graphs appear to go to $\pm\infty$ slowly. In summary, the data support Conjectures 1.2 and 1.3 quite well, possibly with $\delta = 1$ in Conjecture 1.3. It may be that in order for $\mathcal{A}_r(p, X; \phi)$ to tend to a nonzero constant one needs to make a more complicated choice of $\phi(N)$, which potentially depends on r.

We expect these conjectures are fairly robust with respect to the choice of family. E.g., the conjectures should be unchanged if one looks at isomorphism rather than isogeny classes of elliptic curves ordered by conductor (the Stein–Watkins databases also include isomorphism classes), or if one restricts to squarefree conductors. We have also computed some weighted averages in the family $\mathcal{E}_{\mathrm{ht}}$ where one weights by $\phi(H_E)$ rather than $\phi(N_E)$, and the general behavior appears similar. For brevity, we have not included details of weighting by height.

It is well known that one needs to compute very far out to get convincing emperical evidence for the minimalist conjecture, i.e., that the average rank of elliptic curves tends to 0.5. Indeed, the average rank is numerically increasing in the Stein–Watkins database for general conductors—see [BMSW07]. However, in the Stein–Watkins prime conductor database, we see that the average rank per isogeny class quickly goes up to just over 0.98, and then gradually decreases to approximately 0.96544. Moreover, in [BHK$^+$16], the authors find that the average rank increases to about 0.908 around height $6 \cdot 10^8$, and then decreases to around 0.901 by height $2.7 \cdot 10^{10}$.

Given this, it is natural to wonder how much one can trust that our calculations are representative of asymptotic behavior. First, since we are separating by rank, there is no direct effect of the slow convergence to the minimalist conjecture on our data (except that it means we have many curves of rank ≥ 2 in our databases, which is actually helpful for our experiments). It is of course possible that some of our graphs which appear to have a nonzero (or infinite) limit, which would signify a persistent bias, eventually tend to 0, or vice versa. However, we find the asymptotics of the graphs quite compelling up to a factor of order $(1 + O(X^\epsilon))$.

We also note that many other statistics of elliptic curves converge to expectations rather quickly. For instance, in the horizontal direction, the a_p's tend to the Sato–Tate distributions quite quickly. In the vertical direction, numerical convergence to the

parity principle—the notion that half of all curves should have root number $+1$ and half have -1—is also quite fast (e.g., see [BMSW07]).

Consequently we find the computational evidence very convincing of the existence of a rank bias on the order of $O(N^{-\varepsilon})$, even if it is difficult to tease out the exact order of bias from our calculations.

Last, we make a comparison with a naive random model. Suppose the $a_p(E)$'s behave like independent identically distributed random variables. The number of elliptic curves of conductor or discriminant $< X$ is expected to grow like $X^{\frac{5}{6}}$ (see [Wat08]). Hence the variance of sums as in (1.2) should be roughly of order $X^{2\alpha - \frac{5}{6}}$ for $\phi(N) = N^\alpha$ if there is no significant correlation between the rank and $a_p(E)$. In particular, this variance would tend to zero if $\phi(N) \ll N^{\frac{5}{12}-\epsilon}$, which does not agree with the numerical evidence for Conjecture 1.3, and thus even the first statement in Conjecture 1.3(2) supports the notion that the rank of E is correlated with $a_p(E)$.

We remark that similarly modeling $a_p(f)$'s as independent random variables says that the sums in (1.1) should have variance on the order of $X^{2(\alpha-1)}$ for $\phi(N) = N^\alpha$ if there is no significant correlation between the root number and $a_p(f)$. This would suggest that the sums tend to 0 if $\phi(N) \ll N^{1-\epsilon}$, which also does not match Proposition 1.1. (Of course, since most $a_p(f)$ should be Galois conjugates for newform f in a given Atkin-Lehner eigenspace of $S_2(N)$, they are not actually independent.)

2. Traces of Hecke operators

Here we exhibit a bias in the traces of Hecke operators T_n on spaces of newforms with fixed root numbers. Note that in [Mar18, Section 2], explicit dimension formulas for these spaces were proven, and a strict bias towards root number $+1$ was exhibited. That may be viewed as the $n = 1$ analogue of what we do here.

First we set our notation. Denote by $\omega(N)$ the number of prime divisors of N, by $\sigma_1(n)$ the sum-of-divisors function, and by $\delta_{i,j}$ the Kronecker δ function. Let $H(n)$ be the Hurwitz class number, i.e., the number of $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of positive definite integral binary quadratic forms Q of discriminant $-n$ weighted by $\#\mathrm{Aut}(Q)/2$. Throughout, multiple occurrences of $\pm$ and/or $\mp$ within a single statement are to be interpreted as cases dependent on the first occurrence.

Let $S_k(N)$ denote the space of holomorphic even weight k cusp forms for $\Gamma_0(N)$, and $S_k^{\mathrm{new}}(N)$ the subspace spanned by newforms. Let $W_N = \prod_{p|N} W_p$, where W_p denotes the p-th Atkin–Lehner operator on $S_k(N)$. For $S = S_k(N)$ or $S = S_k^{\mathrm{new}}(N)$, denote by $S^\pm$ the subspace of S spanned by eigenforms with W_N-eigenvalue equal to $\pm(-1)^{\frac{k}{2}}$. Then $S_k^{\mathrm{new}}(N)^\pm$ is the space spanned by the newforms with root number ± 1. For a subspace S of $S_k(N)$ and an operator T on $S_k(N)$ which leaves S invariant, denote by $\mathrm{tr}_S T$ the trace of the restriction of T to S.

Lemma 2.1. *Suppose N is squarefree, and $(n, N) = 1$. Then $\mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n W_N = \mathrm{tr}_{S_k(N)} T_n W_N$.*

Proof. This is a special case of [Yam73, Proposition 2]. Here is an alternative argument in terms of representations, which we find more enlightening.

Suppose $f \in S_k^{\mathrm{new}}(M)$ is a newform, where M is a proper divisor of N. Let $\pi = \bigotimes \pi_v$ be the associated cuspidal representation of GL(2), and $\phi = \bigotimes \phi_v$ be a newvector associated to f. Then the contribution $\pi^{K_0(N)}$ of π to $S_k(N)$ has a basis of the form $W_Q \phi$ where Q ranges over positive divisors of M, and $W_Q \phi = \bigotimes \phi_v'$ where $\phi_q' = \pi_q \begin{pmatrix} & 1 \\ q & \end{pmatrix} \phi_q$ if $q|Q$ and $\pi_v' = \pi_v$ if $v \nmid Q$. Now observe that when $(n, N) = 1$, T_n acts by a scalar on $\pi^{K_0(N)}$ and W_N acts as an involution on the above basis elements with no fixed points. Thus $T_n W_N$ has trace zero on $\pi^{K_0(N)}$, and consequently on the whole old space of $S_k(N)$. $\qquad \square$

Proposition 2.2. *Suppose N is squarefree, $n > 1$ is nonsquare, $(n, N) = 1$ and $N > 4n$. Then*

$$\left| \mathrm{tr}_{S_k^{\mathrm{new}}(N)^\pm} T_n \mp \frac{1}{4}(-n)^{\frac{k-2}{2}} H(4nN) \right| < \left(2^{\omega(N)}(4n)^{\frac{k}{2}} + \delta_{k,2} \right) \sigma_1(n).$$

We remark the restriction to n being nonsquare here is purely for simplicity. Otherwise there is an extra term in the explicit formula for $\mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n$ arising in the proof below.

Proof. Let n be a positive integer coprime to N. By definition of our $\pm$-spaces, we have

$$(2.1) \qquad \mathrm{tr}_{S_k^{\mathrm{new}}(N)^\pm} T_n = \frac{1}{2} \left(\mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n \pm (-1)^{\frac{k}{2}} \mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n W_N \right).$$

Yamauchi [Yam73] proved a formula for $\mathrm{tr}_{S_k(N)} T_n W_N$ for general N, though that work contained clerical errors. A corrected form was given by Skoruppa and Zagier [SZ88, (2.7)], which for squarefree N simplifies to:

$$\mathrm{tr}_{S_k(N)} T_n W_N = -\frac{1}{2} \sum_{s^2 \leq 4nN, N|s} p_k(s/\sqrt{N}, n) H(4nN - s^2) + \delta_{k,2} \sigma_1(n).$$

Here, when $b^2 - 4c \neq 0$, $p_k(b, c) = (\rho_1^{k-1} - \rho_2^{k-1})/(\rho_1 - \rho_2)$ where ρ_1, ρ_2 are the roots of $x^2 - bx + c$. If $N > 4n$, we only get the $s = 0$ term in the first sum:

$$(2.2) \qquad \mathrm{tr}_{S_k(N)} T_n W_N = -\frac{1}{2} n^{\frac{k-2}{2}} H(4nN) + \delta_{k,2} \sigma_1(n).$$

By the above lemma, we now have an explicit formula for $\mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n W_N$.

An explicit formula for $\mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n$ is given in [MS10, Theorem 5] for arbitrary N. When N is squarefree and $n > 1$ is nonsquare, this gives

$$\mathrm{tr}_{S_k^{\mathrm{new}}(N)} T_n = -\frac{1}{2} \sum_{t^2 < 4n} p_k(t, n) \sum_{f^2|(4n-t^2)} h_w\left(\frac{t^2 - 4n}{f^2}\right) B_2(N)_f + \delta_{k,2} \mu(N) \sigma_1(n).$$

Here $t \in \mathbb{Z}$, $f \in \mathbb{N}$, $h_w(D)$ is the class number of the imaginary quadratic order $\mathcal{O}(D)$ of discriminant D times $[\mathcal{O}(D)^\times : \mathbb{Z}^\times]$ (interpreted as 0 if D is not a discriminant), and $B_2(N)_f = \prod_{p|N} B_2(p)_f$ where $B_2(p)_f$ is $p - 1$ if $p|f$ and $\left(\frac{t^2 - 4n}{p}\right) - 1$ otherwise.

We note that $\sum_{f^2|(4n-t^2)} h_w\left(\frac{t^2-4n}{f^2}\right) = H(4n - t^2)$ and and $\sum_{t^2 < 4n} H(4n - t^2) < 2\sigma_1(n) - 1$ (e.g., see [MS10, Proposition 12]). Since also $|p_k(t, n)| < 2(4n)^{\frac{k-1}{2}}$ and

8

$B_2(N)_f \leq f2^{\omega(N)}$, we have

$$(2.3) \qquad \left| \text{tr}_{S_k^{\text{new}}(N)} T_n \right| < \left(2^{\omega(N)+1} (4n)^{\frac{k}{2}} + \delta_{k,2} \right) \sigma_1(n).$$

(See also [MS10, Proposition 14] for a similar bound.)

Now combine (2.1) with (2.2) and (2.3). □

In fact one can use the formulas in [Yam73], [SZ88] and [MS10] to give explicit formulas for $\text{tr}_{S_k^{\text{new}}(N)^\pm} T_n$ without assuming N is squarefree (but still coprime to n) or larger than $4n$. However, such formulas will involve alternating sums of class numbers, which makes it more difficult to generalize the following corollary to N non-squarefree.

Corollary 2.3. *Fix $k \geq 2$ even, $n > 1$ squarefree, and $\epsilon > 0$. As $N \to \infty$ along a sequence of squarefree numbers coprime to N, we have*

$$N^{\frac{1}{2}-\epsilon} \ll \pm(-1)^{\frac{k-2}{2}} \, \text{tr}_{S_k^{\text{new}}(N)^\pm} T_n \ll N^{\frac{1}{2}} \log N.$$

Proof. Note that, along a sequence of squarefree N, $2^{\omega(N)} = O(N^\epsilon)$; see for instance the proof of [MW20, Proposition 3.10(ii)]. Consequently,

$$\text{tr}_{S_k^{\text{new}}(N)^\pm} T_n = \pm \frac{1}{4} (-n)^{\frac{k-2}{2}} H(4nN) + O(N^\epsilon).$$

Since nN is squarefree, $4nN$ is a fundamental discriminant and $H(4nN)$ is the usual class number $h(-4nN)$. Now for fundamental discriminants $-D < 0$, we use the standard upper bound $h(-D) \ll D^{\frac{1}{2}} \log D$ and Siegel's (ineffective) lower bound $h(-D) \gg D^{\frac{1}{2}-\epsilon}$. □

Roughly, this says that $\text{tr}_{S_k^{\text{new}}(N)^\pm} T_p$ grows approximately like $\pm\sqrt{N}$ when $k \equiv 2 \bmod 4$, and approximately like $\mp\sqrt{N}$ when $k \equiv 0 \bmod 4$. In particular, when $k = 2$ and ϕ is any weight function, we have that $\mathcal{A}^+(p, X; \phi) > 0$ and $\mathcal{A}^-(p, X; \phi) < 0$ for X sufficiently large. This proves the first part of Proposition 1.1.

Moreover, by [Mar18, Section 2] we have $\dim S_2^{\text{new}}(N)^\pm = \frac{\varphi(N)}{12} + O(N^{\frac{1}{2}} \log N)$. Here φ denotes the Euler totient, not to be confused with a weight function ϕ. Now note that, for any $\epsilon > 0$, we have $X^{2-\epsilon} \ll \sum_{N<X} \varphi(N) \ll X^2$, where in the sum N is restricted to positive squarefree integers. For the lower bound, we are using that the squarefree integers have positive natural density in $\mathbb{N}$ and satisfy $\varphi(N) \geq \frac{N}{2^{\omega(N)}}$, together with the abovementioned fact that $2^{\omega(N)} = O(N^\epsilon)$. Consequently, we see that

$$\frac{\sum_{N<X} N^{\frac{1}{2}-\epsilon} \phi(N)}{X^2} \ll |\mathcal{A}^\pm(p, X; \phi)| \ll \frac{\sum_{N<X} N^{\frac{1}{2}} \log N \phi(N)}{X^{2-\epsilon}}$$

for any weight function ϕ and constant $\epsilon > 0$. (In both sums, N is restricted to positive squarefree N.) This immediately gives the remainder of Proposition 1.1.

3. DATA

Now we present and briefly discuss some data supporting our conjectures.

We computed weighted averages as in (1.2) for the finite subfamilies $\mathcal{E}_{\text{pr}}^{\text{SW}}$, $\mathcal{E}_{\text{all}}^{\text{SW}}$ and $\mathcal{E}_{\text{ht}}^{\text{db}}$ of $\mathcal{E}_{\text{pr}}$, $\mathcal{E}_{\text{all}}$ and $\mathcal{E}_{\text{ht}}$ which respectively consist of all classes of curves contained within
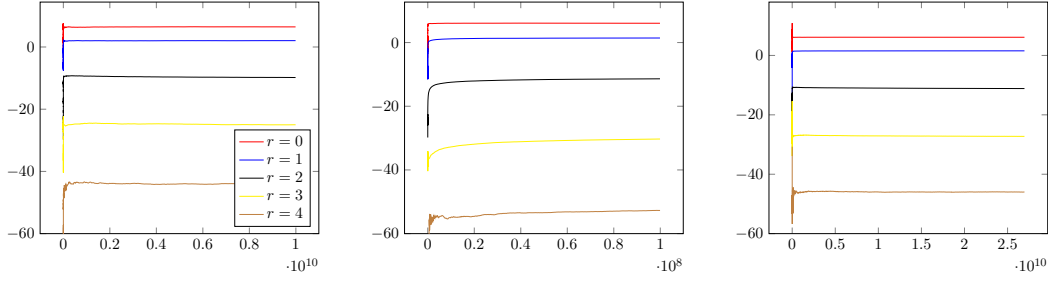
FIGURE 1. Log weight with $p = 7$ and various ranks for $\mathcal{E}_{\text{pr}}^{\text{SW}}$ (left), $\mathcal{E}_{\text{all}}^{\text{SW}}$ (middle), and $\mathcal{E}_{\text{ht}}^{\text{db}}$ (right); all graphs have the same legend
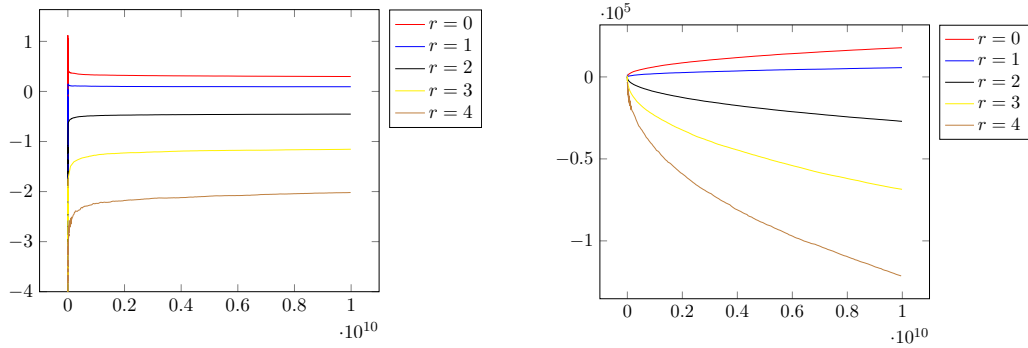

FIGURE 2. Constant (left) and $\sqrt{N}$ (right) weights for $p = 7$ and $\mathcal{E} = \mathcal{E}_{\text{pr}}^{\text{SW}}$

the Stein–Watkins prime conductor database, Stein–Watkins arbitrary conductor database and the height database from [BHK$^+$16]. These calculations only approximate the averages in (1.2) for $X < 10^{10}$, $X \leq 10^8$ and $X \leq 2.7 \times 10^{10}$ for the two reasons mentioned in Section 1: the Stein–Watkins databases are incomplete and many of the ranks in the databases are conjectural.

Assuming correctness of the ranks, we computed the weighted averages $\mathcal{A}_r^{\mathcal{E}}(p, X; \phi)$ for $\mathcal{E} = \mathcal{E}_{\text{pr}}^{\text{SW}}$, $\mathcal{E}_{\text{all}}^{\text{SW}}$ and $\mathcal{E}_{\text{ht}}^{\text{db}}$ for a wide variety of primes $p \leq 300$ and weight functions ϕ for X up to the relevant database bound. On a single CPU core, running calculations for several p at a time, the calculations for a given ϕ took approximately 75–100 minutes of real time for $\mathcal{E}_{\text{pr}}^{\text{SW}}$, 13–16 hours for $\mathcal{E}_{\text{all}}^{\text{SW}}$, and 24–28 hours for $\mathcal{E}_{\text{ht}}^{\text{db}}$.

3.1. **Varying the weight function ϕ.** For a fixed $r \leq 4$ and ϕ, the behavior of $\mathcal{A}_r^{\mathcal{E}}(p, X; \phi)$ is generally similar for both different p and $\mathcal{E}$. See Fig. 1 for overlaid graphs with $p = 7$, $\phi(N) = \log N$, and various r for each $\mathcal{E}$. Subsequent graphs for a variety of p will indicate that the general shapes of the graphs are independent of p, so there is not much lost in specializing these graphs to $p = 7$.

These graphs strongly support Conjecture 1.2, at least for $\phi(N) = \log N$. We have examined similar graphs for a variety of weight functions ϕ and primes p, and these graphs are equally convincing in support of Conjecture 1.2. E.g., see Figs. 2 and 3 for the case $\mathcal{E} = \mathcal{E}_{\text{pr}}^{\text{SW}}$, $p = 7$ with the weights $\phi(N) = 1$, $\phi(N) = \sqrt{N}$, $\phi(N) = N$ and $\phi(N) = N^2$.
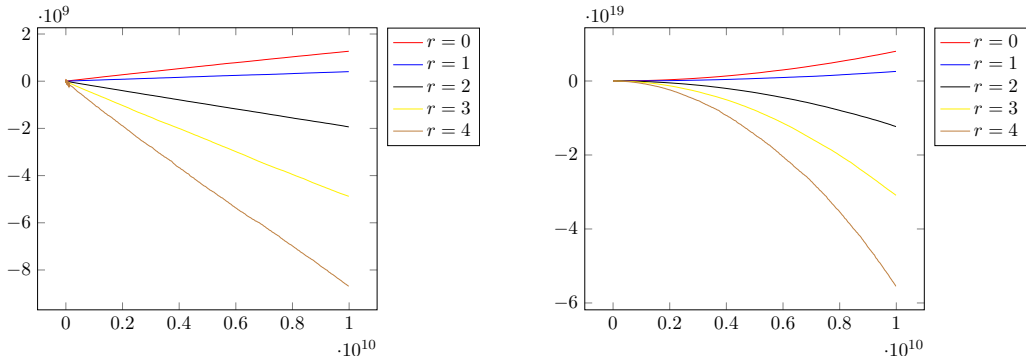
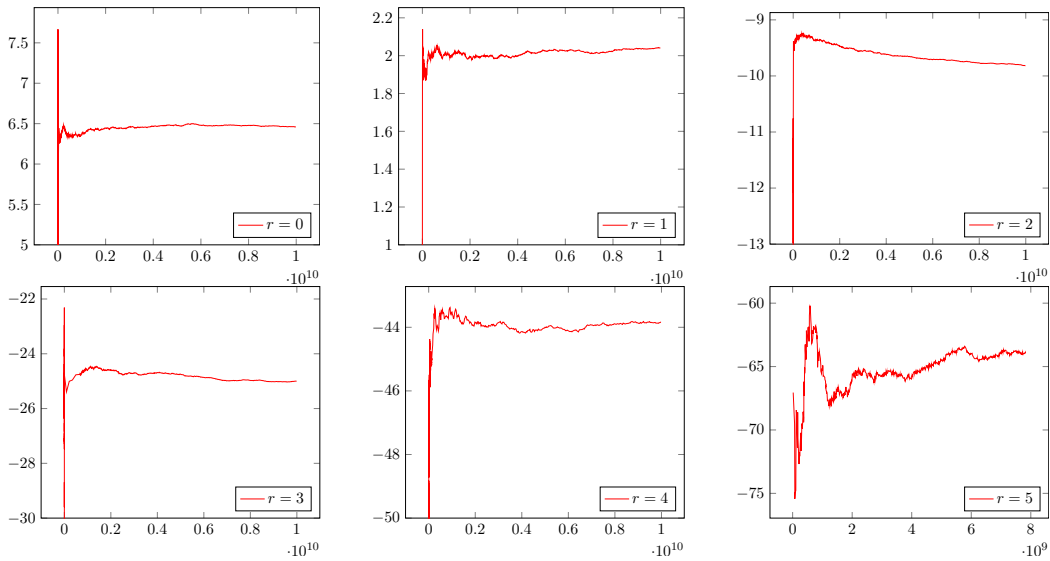FIGURE 3. N (left) and N^2 (right) weights for $p = 7$ and $\mathcal{E} = \mathcal{E}_{\mathrm{pr}}^{\mathrm{SW}}$

FIGURE 4. Log weight for $\mathcal{E}_{\mathrm{pr}}^{\mathrm{SW}}$ with $p = 7$

3.2. **Varying ranks and families for log weighting.** The graphs in Fig. 1 support Conjecture 1.3, but due to the scale the overlaid graphs mask the precise behavior of each rank. See Figs. 4 to 6 for graphs of individual ranks with $p = 7$ and $\phi(N) = \log N$ for the families $\mathcal{E} = \mathcal{E}_{\mathrm{pr}}^{\mathrm{SW}}$, $\mathcal{E}_{\mathrm{all}}^{\mathrm{SW}}$ and $\mathcal{E}_{\mathrm{ht}}^{\mathrm{db}}$, respectively.

Note that for the prime conductor and height databases with log weighting, for each r the graph appears to be eventually relatively flat or possibly tending very slowly away from 0. The case of $\mathcal{E} = \mathcal{E}_{\mathrm{all}}^{\mathrm{SW}}$ is different however: except for $r = 1$ where the graph appears to be slowing increasing away from 0, but for all other ranks the graphs appear to bend towards 0. While it may be that the behavior is actually different for the family $\mathcal{E}_{\mathrm{all}}$, we suspect this difference is more likely due to factors such as the Stein–Watkins all conductor database being rather incomplete (and perhaps giving a biased sample of $\mathcal{E}_{\mathrm{all}}$) and only going up to conductor 10^8. However, even just restricting to the prime conductor and height databases, we see that for some ranks the log weighted averages appear to flatten out and for some ranks (notably $r = 2$ and $r = 3$) they appear to
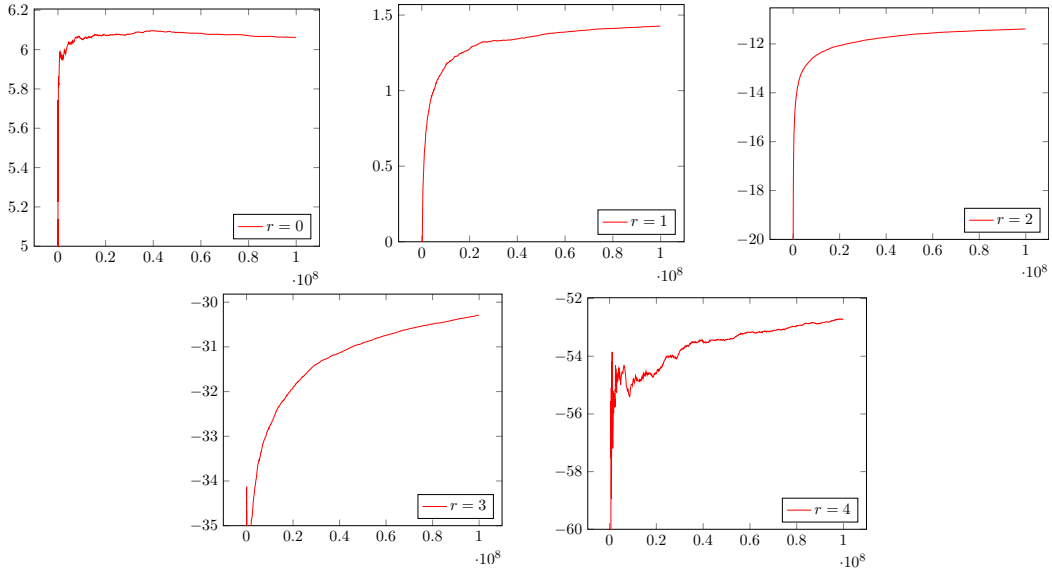
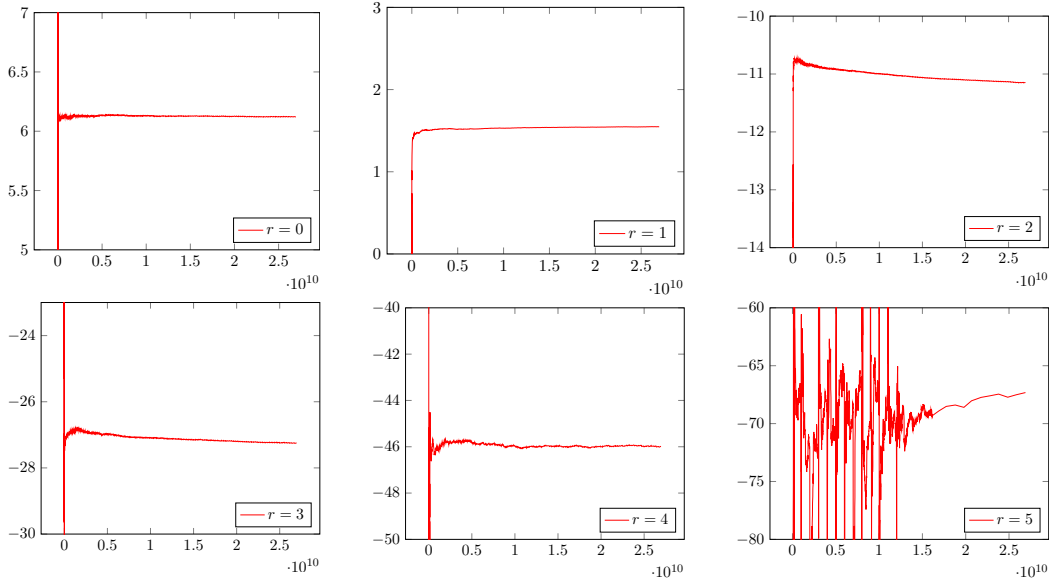FIGURE 5. Log weight for $\mathcal{E}_{\text{all}}^{\text{SW}}$ with $p = 7$

FIGURE 6. Log weight for $\mathcal{E}_{\text{ht}}^{\text{d}}$ with $p = 7$

be slowly tending away from 0. (We have also included $r = 5$ graphs for the prime conductor and height databases, but there is perhaps not enough data to read too much into these graphs.)

3.3. **Log log and polylog weights.** To try to get a more precise sense of the order of bias, in Figs. 7 to 9 we present several graphs for ranks 0–2 for each of our 3 databases. (Analogous graphs for ranks 3 and 4 look similar to the rank 2 graphs, and we omit
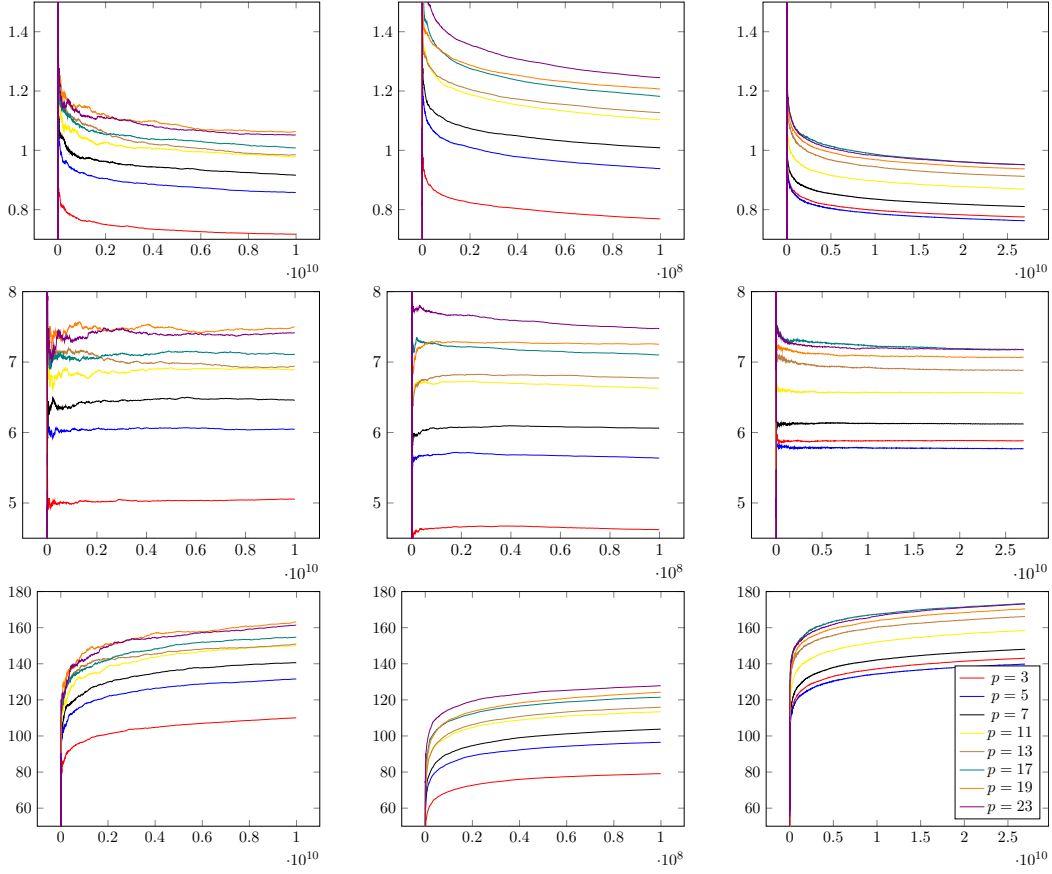
12

FIGURE 7. Rank 0 graphs with log log (top), log (middle) and $\log^2$ (bottom) weights for $\mathcal{E}_{\text{pr}}^{\text{SW}}$ (left) and $\mathcal{E}_{\text{all}}^{\text{SW}}$ (middle), and $\mathcal{E}_{\text{ht}}^{\text{db}}$ (right); all graphs have the same legend

them.) Namely, we graph $\mathcal{A}^{\mathcal{E}}(p, X; \phi)$ for odd $p \leq 23$ with the 3 weight functions $\phi(N) = \log \log N$, $\phi(N) = \log N$ and $\phi(N) = (\log N)^2$.

The first observation is that, indeed, the graphs of $\mathcal{A}^{\mathcal{E}}(p, X; \phi)$ have the same shape for each p. Second, in all cases except for $\mathcal{E} = \mathcal{E}_{\text{all}}^{\text{SW}}$ with $r = 1$ and $\phi(N) = \log \log N$ (which may be different due to limited data as mentioned in the previous subsection), we see that the graphs are tending toward 0 for the log log weight, and tending toward $\pm\infty$ for the $\log^2$ weight, supporting Conjecture 1.3, possibly with $\delta = 1$.

Note that when $r = 0$ and $r = 1$, the log weighted graphs for $\mathcal{E} = \mathcal{E}_{\text{pr}}^{\text{SW}}$ appear to be slightly increasing for some p, roughly flat for other p, and slightly decreasing for other p. Again we suspect these differences are due to limited data, as the prime conductor database contains significantly fewer curves than the other databases.

3.4. **Testing larger p with log weighting.** Our final graphs are intended to illustrate this idea that graphs with limited data may behave differently, especially for larger p. First, consider Fig. 10 which consists of rank 2 graphs for $\mathcal{E} = \mathcal{E}_{\text{all}}^{\text{SW}}$ and log weight for a variety of primes $p \leq 271$. Note these graphs all have the same shape. (Of course we expect that the weighted averages tend to have larger absolute value for larger p
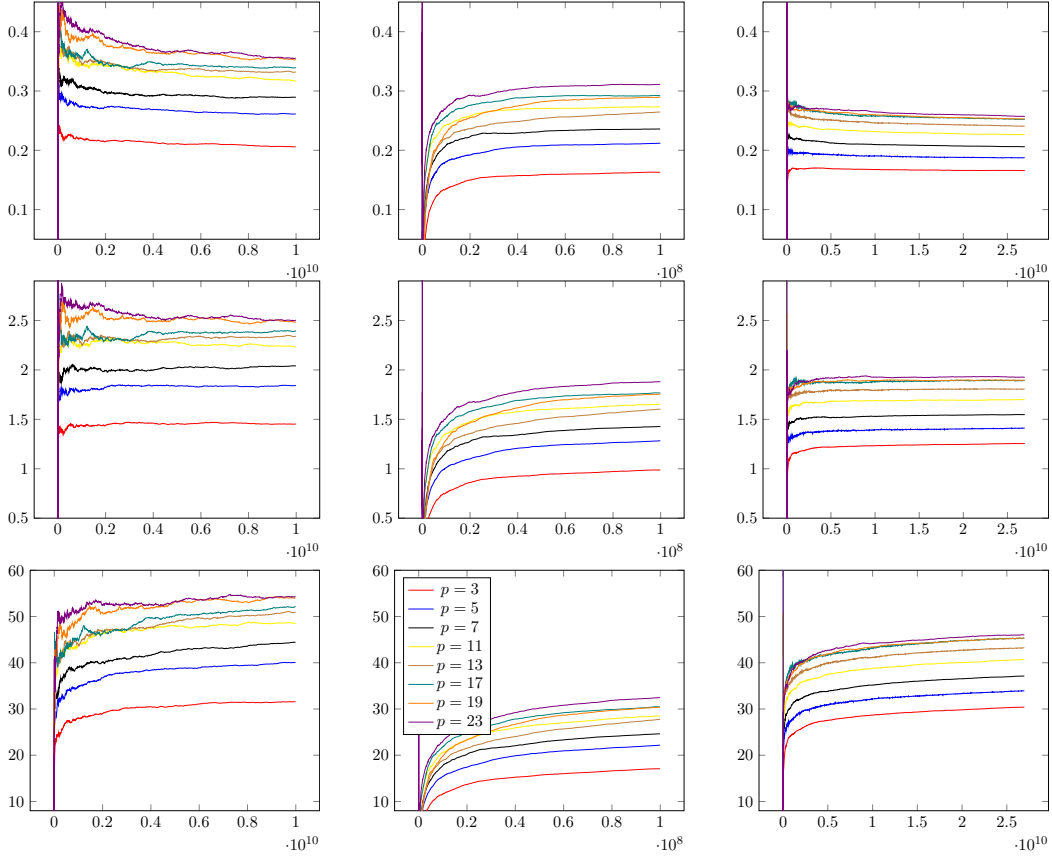
13

FIGURE 8. Rank 1 graphs with $\log\log$ (top), $\log$ (middle) and $\log^2$ (bottom) weights for $\mathcal{E}_{\mathrm{pr}}^{\mathrm{SW}}$ (left) and $\mathcal{E}_{\mathrm{all}}^{\mathrm{SW}}$ (middle), and $\mathcal{E}_{\mathrm{ht}}^{\mathrm{db}}$ (right); all graphs have the same legend

because the Hecke bound grows with p.) If one plots the same graphs for rank 0 or 1, we again see the shape is essentially independent of p (the case $p = 79$ case with $r = 0$ is an exception).

However, if we look at rank 3 graphs the shape does seem to depend on p for the larger values of p using the Stein–Watkins database, but not the height database. See Fig. 11. We attribute this to there being fewer rank 3 curves in the databases than ranks 0, 1 or 2, but there being many more curves in the height database than the Stein–Watkins database.

Thus we expect that the general behavior of weighted averages is independent of p, but it may require a very large amount of data to numerically see the asymptotic behavior, especially for larger p and ϕ having logarithmic growth.

3.5. **Summary.** In summary, we find the data quite suggestive that, for each of our 3 families, the weighted averages $\mathcal{A}_r^{\mathcal{E}}(p, X; \phi)$ tend to 0 if $\phi(N) \ll \log\log N$ and tend to $+\infty$ (resp. $-\infty$) if $\phi(N) \gg (\log N)^2$ and $r \leq 1$ (resp. $r \geq 2$). When $\phi(N) = \log N$, some of the graphs appear very flat, and some do not. Because of the variation of these
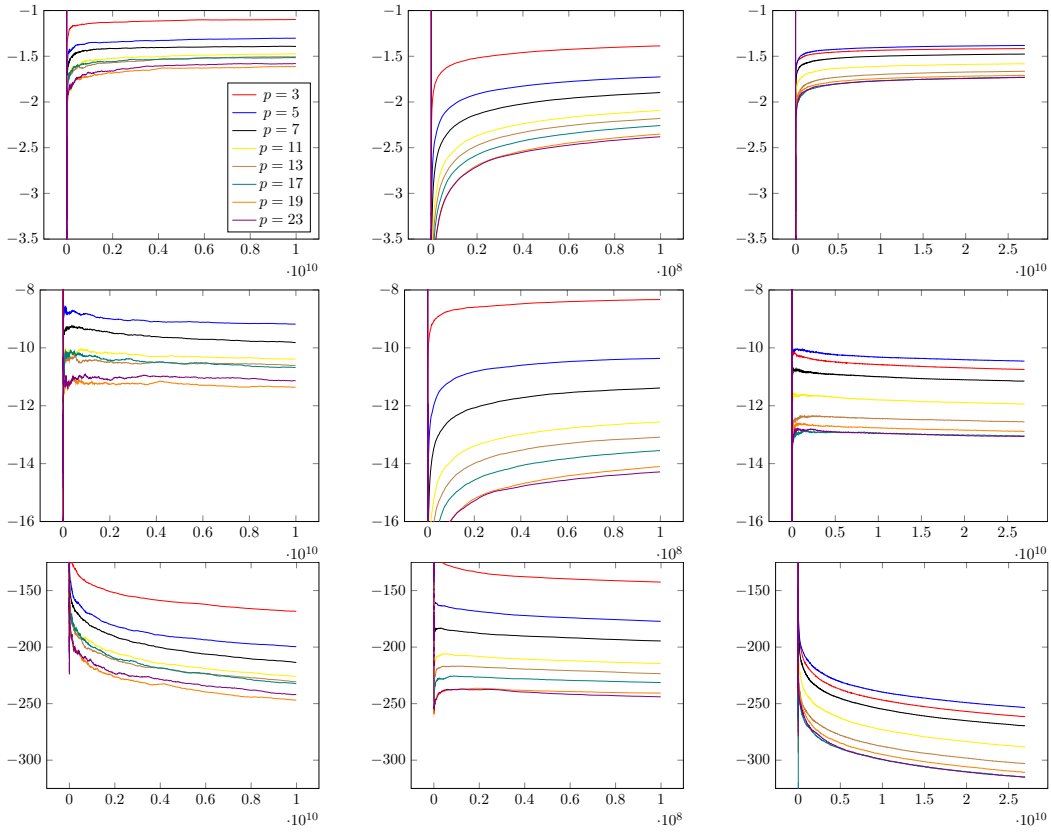
14

FIGURE 9. Rank 2 graphs with log log (top), log (middle) and $\log^2$ (bottom) weights for $\mathcal{E}_{\mathrm{pr}}^{\mathrm{SW}}$ (left) and $\mathcal{E}_{\mathrm{all}}^{\mathrm{SW}}$ (middle), and $\mathcal{E}_{\mathrm{ht}}^{\mathrm{db}}$ (right); all graphs have the same legend
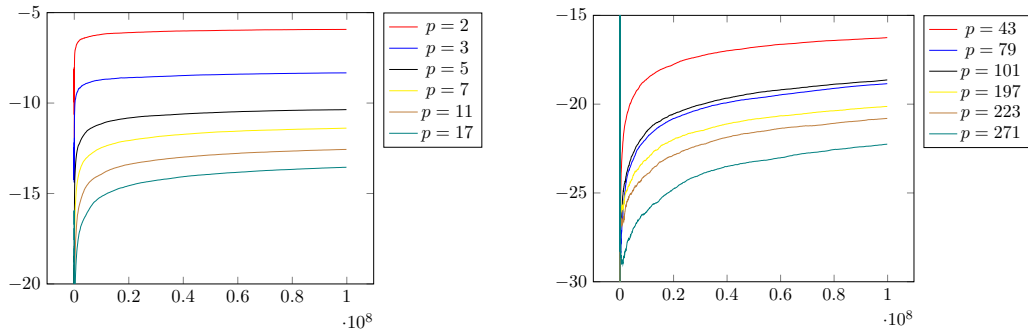

FIGURE 10. Log weight with $r = 2$ and various p for $\mathcal{E}_{\mathrm{all}}^{\mathrm{SW}}$

graphs, the precise asymptotic behavior is not apparent with the databases currently available to us, but it seems plausible that the value of δ in Conjecture 1.3 may be 1.
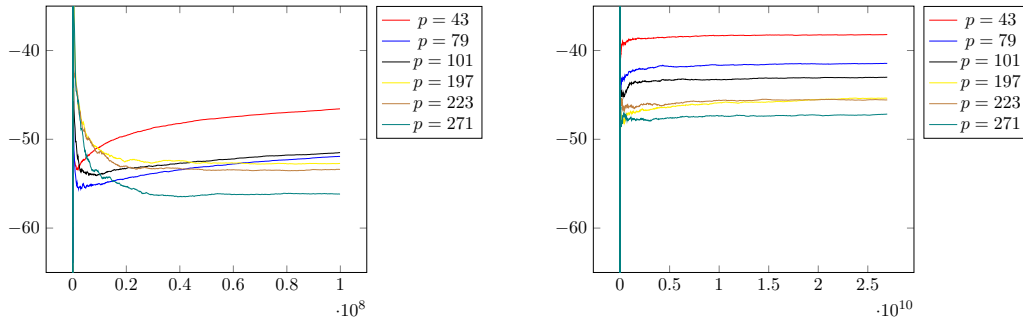
15

FIGURE 11. Log weight with $r = 3$ and various p for $\mathcal{E}_{\text{all}}^{\text{SW}}$ (left) and $\mathcal{E}_{\text{ht}}^{\text{db}}$ (right)

REFERENCES

[ACF+]   Megumi Asada, Ryan C. Chen, Eva Fourakis, Yujin Hong Kim, Andrew Kwon, Jared Duker Lichtman, Blake Mackall, Steven J. Miller, Eric Winsor, Karl Winsor, Jianing Yang, and Kevin Yang, *Lower-Order Biases in the Second Moment of Dirichlet Coefficients in Families of L-Functions*. Experiment. Math., to appear.

[BHK+16] Jennifer S. Balakrishnan, Wei Ho, Nathan Kaplan, Simon Spicer, William Stein, and James Weigandt, *Databases of elliptic curves ordered by height and distributions of Selmer groups and ranks*, LMS J. Comput. Math. **19** (2016), no. suppl. A, 351–370.

[BLGHT11] Tom Barnet-Lamb, David Geraghty, Michael Harris, and Richard Taylor, *A family of Calabi-Yau varieties and potential automorphy II*, Publ. Res. Inst. Math. Sci. **47** (2011), no. 1, 29–98, DOI 10.2977/PRIMS/31. MR2827723

[BMSW07] Baur Bektemirov, Barry Mazur, William Stein, and Mark Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 2, 233–254.

[BGR19]   Michael A. Bennett, Adela Gherga, and Andrew Rechnitzer, *Computing elliptic curves over* $\mathbb{Q}$, Math. Comp. **88** (2019), no. 317, 1341–1390.

[Bir68]   B. J. Birch, *How the number of points of an elliptic curve over a fixed prime field varies*, J. London Math. Soc. **43** (1968), 57–60.

[BSD65]   B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves. II*, J. Reine Angew. Math. **218** (1965), 79–108.

[BJ16]   Dongho Byeon and Keunyoung Jeong, *Infinitely many elliptic curves of rank exactly two*, Proc. Japan Acad. Ser. A Math. Sci. **92** (2016), no. 5, 64–66.

[KN]   Matija Kazalick and Bartosz Naskrecki, *Second moments and the bias conjecture for the family of cubic pencils*. arXiv:2012.11306.

[MMRW16] Blake Mackall, Steven J. Miller, Christina Rapti, and Karl Winsor, *Lower-order biases in elliptic curve Fourier coefficients in families*, Frobenius distributions: Lang-Trotter and Sato-Tate conjectures, Contemp. Math., vol. 663, Amer. Math. Soc., Providence, RI, 2016, pp. 223–238.

[Mar18]   Kimball Martin, *Refined dimensions of cusp forms, and equidistribution and bias of signs*, J. Number Theory **188** (2018), 1–17.

[Mar21]   ———, *An on-average Maeda-type conjecture in the level aspect*, Proc. Amer. Math. Soc. **149** (2021), 1373–1386.

[MW20]   Kimball Martin and Jordan Wiebe, *Zeroes of quaternionic modular forms and central L-values*, J. Number Theory **217** (2020), 460–494.

[MS10]   M. Ram Murty and Kaneenika Sinha, *Factoring newparts of Jacobians of certain modular curves*, Proc. Amer. Math. Soc. **138** (2010), no. 10, 3481–3494.

[Nag97] Koh-ichi Nagao, $\mathbf{Q}(T)$-*rank of elliptic curves and certain limit coming from the local points*, Manuscripta Math. **92** (1997), no. 1, 13–32. With an appendix by Nobuhiko Ishida, Tsuneo Ishikawa and the author.

[PPVW19] Jennifer Park, Bjorn Poonen, John Voight, and Melanie Matchett Wood, *A heuristic for boundedness of ranks of elliptic curves*, J. Eur. Math. Soc. (JEMS) **21** (2019), no. 9, 2859–2903.

[RS98] Michael Rosen and Joseph H. Silverman, *On the rank of an elliptic surface*, Invent. Math. **133** (1998), no. 1, 43–67.

[SZ88] Nils-Peter Skoruppa and Don Zagier, *Jacobi forms and a certain space of modular forms*, Invent. Math. **94** (1988), no. 1, 113–146.

[SW02] William A. Stein and Mark Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, pp. 267–275.

[Wat08] Mark Watkins, *Some heuristics about elliptic curves*, Experiment. Math. **17** (2008), no. 1, 105–125.

[Yam73] Masatoshi Yamauchi, *On the traces of Hecke operators for a normalizer of* $\Gamma_0(N)$, J. Math. Kyoto Univ. **13** (1973), 403–411.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF OKLAHOMA, NORMAN, OK 73019
E-mail address: kimball.martin@ou.edu

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, IN 47405