

MACHINE LEARNING THE ARITHMETIC OF BOYD'S MAHLER MEASURE CONJECTURES

ALBERTO ALFARANO, PABLO BIANUCCI, MATILDE N. LALÍN, AND BEREND RINGELING

ABSTRACT. Boyd conjectured that the Mahler measure of $P_k(x, y) = x + y + \frac{1}{x} + \frac{1}{y} + k$ for k an integer, is given by $r_k L'(E_k, 0)$, where E_k is the elliptic curve associated to the zero locus of P_k and r_k is a rational number. We study various arithmetic properties of r_k using a dataset containing the first 250,000 values of k , combining large-scale statistical analysis assisted by CLAUDE with transformer-based experiments carried out using AXOLVER.

We recover Boyd's observation that, apart from a few exceptions, r_k is the reciprocal of an integer. The size of this integer is governed by the conductor of the elliptic curve. Moreover, its p -adic valuations display markedly different behavior according to the prime. For $p \geq 5$, the probability of $v_p(r_k) = -m$ for $m \geq 1$ appears to be p^{-m} . For the primes 2 and 3, however, we find additional arithmetic structure involving congruence conditions on k and the primes of bad reduction of E_k . Although the neural networks do not predict r_k exactly, they recover significant information about its magnitude and valuations. In particular, the experiments at the prime 2 suggest arithmetic structure beyond the explicit predictor obtained from our statistical analysis.

The (logarithmic) Mahler measure of a nonzero rational function $P \in \mathbb{C}(x_1, \dots, x_n)$ is defined by

$$m(P) = \frac{1}{(2\pi i)^n} \int_{\mathbb{T}^n} \log |P(x_1, \dots, x_n)| \frac{dx_1}{x_1} \dots \frac{dx_n}{x_n},$$

where $\mathbb{T}^n = \{(u_1, \dots, u_n) \in \mathbb{C}^n : |u_1| = \dots = |u_n| = 1\}$. In the one variable polynomial case given by $n = 1$, Jensen's formula leads to an expression in terms of the roots of P . More precisely, if $P(x) = a \prod_j (x - \alpha_j)$, one can see that $m(P) = \log |a| + \sum_j \log \max\{1, |\alpha_j|\}$.

However, for $n > 1$, the nature of the values is more mysterious. In many cases where P has rational coefficients, the Mahler measure has been related to values of functions with arithmetic significance.

Deninger [Den97] predicted a connection of certain two-variable Mahler measures with special values of L functions associated to elliptic curves via Beilinson's conjectures, which are very general frameworks relating special values of L -functions to regulators, generalizing central statements such as the Birch and Swinnerton Dyer conjecture and the Dirichlet class number formula. These connections were verified by Boyd [Boy98] and Rodriguez-Villegas [RV99] in many cases, most of them numerical, with a few formulas proven by Brunault, Mellit, Rogers and others [Bru06, Lal10, LR07, Mel19, RZ12, RZ14, Zud14]. One of the most studied families fitting the above framework is the one proposed by Boyd [Boy98], given by

$$\mu_k := m \left(x + y + \frac{1}{x} + \frac{1}{y} + k \right),$$

where k is an integral parameter. Boyd computed μ_k for $k \in \mathbb{Z} \cap [0, 40]$ and combined these observations with Deninger's predictions to conjecture that for $k \in \mathbb{Z} \setminus \{-4, 0, 4\}$,

$$(0.1) \quad \mu_k = r_k L'(E_k, 0),$$

where E_k is the elliptic curve corresponding to the projective closure of the desingularization of $x + y + \frac{1}{x} + \frac{1}{y} + k = 0$, and r_k is a nonzero rational number. In fact, apart from a few exceptions,

it seems that $1/r_k$ is an integer. Table 0.1 shows $1/r_k$ for the first 40 values of k . The conductor N_k is included in the table, since it determines the L -function. Moreover, the sign of r_k should be determined by the root number ω_k of E_k , as the Mahler measure is positive. We remark, by using the change of variables $x \rightarrow -x$ and $y \rightarrow -y$ in the integral, that $\mu_k = \mu_{-k}$, and therefore it suffices to consider positive k . When $k = 0, 4$, the zero loci of the corresponding Laurent polynomials have genus 0. In these cases, $\mu_0 = 0$ and $\mu_4 = 2L'(\chi_{-4}, -1)$, the Dirichlet series on the character of conductor 4 [Boy98].

k	$1/r_k$	N_k	k	$1/r_k$	N_k
1	1	15	21	-12	1785
2	1	24	22	24	3432
3	1/2	21	23	6	1311
4	-	-	24	8	840
5	1/6	15	25	-16	3045
6	2	120	26	128	17160
7	2	231	27	12	2139
8	1/4	24	28	-2	336
9	2	195	29	-24	4785
10	-8	840	30	240	26520
11	-8	1155	31	-16	3255
12	1/2	48	32	1/3	42
13	-4	663	33	-204	35409
14	8	840	34	256	38760
15	-24	3135	35	224	42315
16	1/11	15	36	2	240
17	-24	4641	37	-208	50061
18	-16	1848	38	288	54264
19	-40	6555	39	336	58695
20	2	240	40	-8	1320

TABLE 0.1. Numerical values of r_k and the conductor of E_k .

Some of these identities have been proven, see Table 0.2.

k	$1/r_k$	Conductor of E_k	Reference
1	1	15	Rogers and Zudilin [RZ14]
2	1	24	Rogers and Zudilin [RZ12]
3	1/2	21	Brunault [Bru16]; Lalín, Samart, and Zudilin [LSZ16]
5	1/6	15	Lalín [Lal10] from $k = 1$
8	1/4	24	Lalín and Rogers [LR07] from $k = 2$
12	1/2	48	Brunault [Bru16]
16	1/11	15	Lalín [Lal10] from $k = 1$

TABLE 0.2. Proven formulas for integer values of k .

Understanding the nature of these identities would shed light on our understanding of special values of L functions and Beilinson's conjectures. The following is a natural question to pose.

Question 0.1. *How does r_k relate to k and to the elliptic curve E_k ?*

We remark that r_k should be predicted by the Bloch–Kato conjectures. Very recent work by Dummigan, Golyshev, de Jeu and Kerr [DGdJK26] investigates the 2-part of the Bloch–Kato conjecture for this family of curves.

We have computed the first 250,000 values of μ_k and explored this question in two different ways: by analyzing the data with the assistance of CLAUDE and by training neural networks on the problem via AXOLVER, with the goal of predicting certain properties of r_k .

The data analysis revealed several arithmetic regularities exhibited by r_k . After accounting for the sign, which is governed by the root number, the reciprocal $n_k = 1/r_k$ is almost always an integer. Its order of magnitude is explained by the conductor through the estimate

$$|n_k| \asymp \frac{N_k}{\log(k)},$$

while its p -adic valuations show different behaviors depending on the prime. For primes $p \geq 5$, the data suggest the simple law

$$\mathbb{P}(v_p(n_k) = r) \stackrel{?}{=} p^{-r} \quad (r \geq 1).$$

The primes 2 and 3, however, have a different behavior. The 3-adic valuation seems to be strongly influenced by congruence conditions on k and by the parity of the number of bad primes congruent to 1 mod 3. The 2-adic valuation appears to be well modeled by a deterministic term depending on primes of bad reduction plus an apparently random negative-binomial error term. More precisely, we observe that

$$v_2(n_k) \stackrel{?}{=} \omega_{\text{odd}}(k) + 2\omega_{\text{odd}}(k^2 - 16) - c(k) + s(k) + X_k,$$

where $\omega_{\text{odd}}(n)$ counts the odd primes dividing n , and $c(k)$ and $s(k)$ are small bounded correction terms that codify the behavior of E_k at 2, defined in (3.8) and (3.10), respectively. Finally, the empirical distribution of the residual X_k is well approximated by the negative-binomial distribution $\text{NB}(r = 3, p = \frac{3}{5})$. We emphasize that this is a heuristic observation rather than a precise conjecture. The work of Dummigan, Golyshev, de Jeu and Kerr [DGdJK26] provides a formula for $v_2(n_k)$ for an infinite, positive-density family of k , conditionally on the Bloch–Kato conjecture. The formula depends on the 2-valuation of a Selmer group associated to an elliptic curve 2-isogenous to E_k as well as on a regulator index $\iota_{k/4,2} \in \{1, 2\}$, which measures the index of an explicit motivic cohomology class in the rank-one $\mathbb{Z}_{(2)}$ -lattice determined by the 2-adic regulator. We briefly examine this connection in Section 3.6.

The experiments showed us that AXOLVER is able to recover part of this arithmetic structure from the data, but that the full problem of predicting r_k remains very difficult. While the model is unable to completely learn r_k , it does learn meaningful information about its size, with the best behavior occurring when the conductor is included among the input features. In particular, the model begins to reproduce the observed distribution of $L(E_k, 2)$. The valuation experiments are more mixed. For $p = 5$ and $p = 7$, the model tends to collapse to the most frequent output $v_p(n_k) = 0$, so these experiments do not yet reveal useful structure. By contrast, the experiments for $v_2(n_k)$ and $v_3(n_k)$ show that the network can exploit arithmetic features such as the factorization of the discriminant or conductor, giving evidence that the statistical patterns observed in Section 3 are visible to a learning model. In the case of the prime 2, the model's learning seems to go beyond the understanding that we can derive from the data (see Section 5.4.2).

This article is organized as follows. Section 1 presents the general mathematical background for the problem. Section 2 briefly discusses the generation of the data. The statistical observations are presented in Section 3. Section 4 describes the model and the experiments run in AXOLVER, while

Section 5 contains a detailed discussion on the results of those experiments. Section 6 summarizes our findings and discusses directions for further research, while Section 7 contains an appendix with supplementary tables and figures.

ACKNOWLEDGMENTS

This project was initiated during the workshop “Using AI to find examples” that took place in April 2026 at the Centre de recherches mathématiques. The authors are grateful to the organizers, particularly Andrew Granville, and to the CRM for the excellent working conditions. The authors are thankful to François Charton, Christopher Deninger, Jordan Ellenberg, Dimitris Koukolopoulos, Shousen Lu, Riccardo Pengo, and Wadim Zudilin for several helpful discussions. Part of this work was conducted during the workshop “AI and number theory” that took place in May 2026 at the American Institute of Mathematics. The authors thank the organizers and Anthropic for providing access and guidance to CLAUDE MAX during this event.

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2019-06988 to PB, RGPIN-2022-03651 to ML), and the Fonds de recherche du Québec - Nature et technologies (Projet de recherche en équipe 345672 to ML), the Institut des sciences mathématiques (postdoctoral fellowship to BR), the Centre de recherches mathématiques (postdoctoral fellowship to BR).

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

During the preparation of this manuscript, the authors used artificial intelligence (AI) tools to assist with data analysis, machine learning modeling, and manuscript editing:

- Data Analysis: Anthropic’s CLAUDE model was utilized to process, clean, and identify patterns within the research dataset.
- Machine Learning: Axiom’s AXOLVER framework was used to train sequence-to-sequence neural networks and execute mathematical/symbolic reasoning tasks within the study’s machine learning pipeline.
- Writing and Editing: OpenAI’s CHATGPT was used to improve the readability, grammar, and style of the author-drafted text.

The original research framework, data collection, and final interpretations were entirely conceptualized and directed by the human authors. After using these tools, the authors reviewed, verified, and edited all outputs to ensure accuracy and eliminate potential hallucinations. The authors accept full accountability for the content and integrity of the final work.

DATA AVAILABILITY STATEMENT

The code used in this study is available in the GITHUB repository <https://github.com/BerendRingeling/Mahler>. The datasets are available through the DROPBOX link provided in that repository.

1. BACKGROUND AND GENERALITIES

The plane curve defined by the equation

$$x + \frac{1}{x} + y + \frac{1}{y} + k = 0,$$

where k is a parameter, is birationally equivalent to an elliptic curve (over $\mathbb{Q}(k)$). This can be seen by applying, for example, the birational change of variables

$$x = \frac{8(k^2X + 8Y)}{k(k^2 - 16X)}, \quad y = \frac{8(k^2X - 8Y)}{k(k^2 - 16X)},$$

with inverse

$$X = -\frac{k^2xy}{16}, \quad Y = \frac{k^3(xy + 1)(x - y)}{128},$$

which leads to the Weierstrass equation

$$(1.1) \quad E_k : Y^2 = X \left(X^2 + \frac{k^2}{8} \left(\frac{k^2}{8} - 1 \right) X + \frac{k^4}{16^2} \right).$$

For fixed k , this curve is nonsingular provided that $k \neq 0, \pm 4$.

If an elliptic curve E is defined over $\mathbb{Q}$, one can construct its L -function as follows

$$L(E, s) = \prod_{\text{good } p} (1 - a_p p^{-s} + p^{1-2s})^{-1} \prod_{\text{bad } p} (1 - a_p p^{-s})^{-1} = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

where for p prime,

$$a_p = 1 + p - \#E(\mathbb{F}_p)$$

denotes the trace of Frobenius. Here the bad p are the primes where the corresponding reduction of E modulo p is singular, and the good p are all the other primes. The bad primes are exactly those that divide the conductor N of the elliptic curve.

The Weierstrass model (1.1) is defined over $\mathbb{Q}(k^2)$. This observation prompted Rodriguez-Villegas [RV99] to conjecture that equation (0.1) extends to the case where k^2 is a positive integer and to prove the cases $k = 2\sqrt{2}, 4\sqrt{2}$. A word of caution here is that, since $k \notin \mathbb{Q}$, and any change of variables will be defined over $\mathbb{Q}(k)$, the choice of the model does make a difference. A few more cases where $k^2 \in \mathbb{Z}$ have been proven, see Table 1.1.

k	$1/r_k$	Conductor of E_k	Reference
$\sqrt{2}$	4	56	Zudilin [Zud14]
$2\sqrt{2}$	1	32	Rodriguez-Villegas [RV99]
$3\sqrt{2}$	2/5	24	Lalín [Lal10] from $k = 2$
$4\sqrt{2}$	1	64	Rodriguez-Villegas [RV99]
i	1/2	17	Zudilin [Zud14]
$2i$	1	40	Mellit [Mel11]
$3i$	1/5	15	Lalín [Lal10] from $k = 1$
$i\sqrt{2}$	2/3	24	Lalín [Lal10] from $k = 2$

TABLE 1.1. Proven formulas for quadratic irrational and non-real values of k .

In the present work, however, we restrict our statistical and machine-learning analysis to integral values of k . This is the setting of Boyd's original numerical experiments. This restriction is also convenient computationally: generating data for $k = 1, \dots, 2B$ appears slightly faster than carrying out the corresponding computations for all integral values $|m| \leq B$, where $m = k^2$ is no longer required to be a square. We briefly examine the nonsquare case in Section 3.5, where the near-integrality, and valuation laws are found to persist in many cases.

The motivation for conjecturing formulas such as (0.1) comes from Beĭlinson's conjectures, as explained in [Boy98, Den97, RV99].

2. NUMERICAL GENERATION OF DATA

In order to generate the data needed to train the model, we need to compute both sides of equation (0.1). The computation of $L'(E_k, 0)$ can be performed to very high precision in PARI/GP [PAR25]. For the Mahler measure we have the following explicit formulas for real k (see [RZ14, p. 2324]):

$$(2.1) \quad \mu_k = \begin{cases} \log |k| - \frac{2}{k^2} {}_4F_3\left(\frac{3}{2}, \frac{3}{2}, 1, 1; 2, 2, 2; \frac{16}{k^2}\right), & \text{if } |k| \geq 4, \\ \frac{|k|}{4} {}_3F_2\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; \frac{3}{2}, 1; \frac{k^2}{16}\right), & \text{if } |k| < 4. \end{cases}$$

Here ${}_4F_3$ and ${}_3F_2$ denote hypergeometric functions. This explicit description allows one to compute μ_k to high precision for $k \in \mathbb{Z} \setminus \{-4, 0, 4\}$. Both μ_k and $L'(E_k, 0)$ were computed using 19 decimal digits of precision.

Using the LLL algorithm (implemented as `linddep` in PARI/GP [PAR25]), we can then quickly numerically recognize the rational quotient r_k .

We computed the values r_k for $1 \leq k \leq 250,000$ (with $k \neq 4$). In total this required about 9.6×10^7 s (≈ 3 years) of single-core-equivalent computation time, with the average computation time per value rising from roughly 35 s for small k to about 16.5 min for k near 250,000. The computations were carried out on 21 CPU cores across 17 servers with a combined 2108 GB (≈ 2.1 TB) of memory.

3. STATISTICAL OBSERVATIONS OF THE DATA

We analyzed the values r_k for $1 \leq k \leq 250,000$. First of all, since the Mahler measure is always positive, the sign of r_k is determined by the sign of $L'(E_k, 0)$, which, by the functional equation of the L -function, is determined by the root number ω_k of E_k . Secondly, for all but a few exceptional values of k in the dataset, namely for

$$(3.1) \quad k \notin \{3, 4, 5, 8, 12, 16, 32\},$$

the quantity r_k is conjecturally the reciprocal of an integer. Inspired by this, we write $n_k = 1/r_k$, so that n_k is an integer for all the k satisfying (3.1). We make the following empirical observations on the size of n_k and its p -adic valuations for primes p for these k .

3.1. The size of $|n_k|$. We can use the approximation for μ_k and $L'(E_k, 0)$ to find an estimate for $|n_k|$. Using the hypergeometric description for the Mahler measure (2.1), we find

$$(3.2) \quad \mu_k = \log(k) + o(1)$$

as $k \rightarrow \infty$. For $L'(E_k, 0)$, we use the functional equation to arrive at

$$(3.3) \quad L'(E_k, 0) = \omega_k \frac{N_k}{4\pi^2} L(E_k, 2),$$

where N_k is the conductor of E_k and ω_k the root number. Using the following lemma, we can estimate the size of $L(E_k, 2)$.

Lemma 3.1. *For any elliptic curve E over $\mathbb{Q}$, we have the bounds*

$$(0.2117 \dots) = \frac{\zeta(3)^2}{\zeta(3/2)^2} \leq L(E, 2) \leq \zeta(3/2)^2 (= 6.8245 \dots).$$

Proof. We write the L -value in terms of its local factors

$$L(E, 2) = \prod_{\text{primes } p} L_p(E, 2),$$

where

$$L_p(E, 2) = \begin{cases} (1 - a_p p^{-2} + p \cdot p^{-4})^{-1}, & p \nmid N \\ (1 \pm p^{-2})^{-1}, & p \parallel N \\ 1, & p^2 \mid N, \end{cases}$$

where N is the conductor of E . For good primes (i.e. $p \nmid N$), we have, by the Hasse-Weil bound, $|a_p| \leq 2\sqrt{p}$, and therefore,

$$(1 - p^{-3/2})^2 = 1 - 2p^{-3/2} + p^{-3} \leq 1 - a_p p^{-2} + p \cdot p^{-4} \leq 1 + 2p^{-3/2} + p^{-3} = (1 + p^{-3/2})^2.$$

Taking reciprocals,

$$\frac{1}{(1 + p^{-3/2})^2} \leq L_p(E, 2) \leq \frac{1}{(1 - p^{-3/2})^2}.$$

For $p \parallel N$, the same estimates hold:

$$(1 - p^{-3/2})^2 \leq 1 - p^{-2} \leq 1 \pm p^{-2} \leq 1 + p^{-2} \leq (1 + p^{-3/2})^2,$$

where the outer inequalities can be easily verified. Finally, for $p^2 \mid N$, it is immediate that these bounds hold. Taking the product over all primes p , we conclude that

$$\frac{\zeta(3)^2}{\zeta(3/2)^2} = \prod_{\text{primes } p} \frac{1}{(1 + p^{-3/2})^2} \leq L(E, 2) \leq \prod_{\text{primes } p} \frac{1}{(1 - p^{-3/2})^2} = \zeta(3/2)^2.$$

□

As a consequence of (3.2), (3.3) and Lemma 3.1, we conclude that

$$(3.4) \quad |n_k| = \frac{|L'(E_k, 0)|}{\mu_k} \asymp \frac{N_k}{\log(k)}.$$

Thus, the order of magnitude of n_k is completely determined by k and the conductor N_k . In Figure 3.1 we see the distribution of $L(E_k, 2)$ in our dataset.

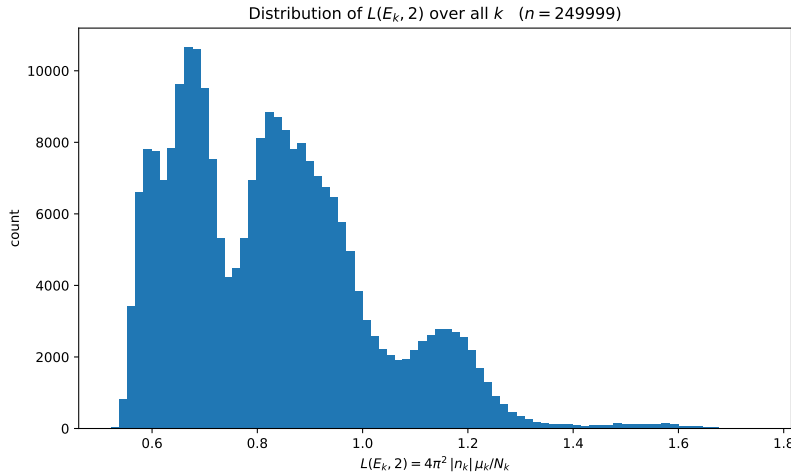


FIGURE 3.1. Values of $L(E_k, 2)$ for $k \leq 250,000$ in our dataset.

Although Lemma 3.1 only forces $(0.2117, 6.8245)$, the values we observe are in a narrower range $(0.5222, 1.7532)$, with mean value 0.8326. The exact distribution of $L(E_k, 2)$ is not known to us; here we record only its empirical behavior and its role in governing the size of n_k through (3.4). We do not assert any direct relation between this distribution and the p -adic valuations of n_k considered below.

3.2. p -adic observations for primes $p \geq 5$. We observe empirically that

$$\mathbb{P}(v_p(n_k) \geq 1) = \frac{1}{p-1},$$

rather than the probability $1/p$ predicted by a naive uniform-distribution model modulo p . Conditional on $p \mid n_k$, however, the higher valuations appear to follow the usual geometric distribution:

$$\mathbb{P}(v_p(n_k) = r \mid v_p(n_k) \geq 1) = \left(1 - \frac{1}{p}\right) p^{-(r-1)}, \quad r \geq 1.$$

Equivalently,

$$\mathbb{P}(v_p(n_k) = r) = \frac{1}{p^r}, \quad r \geq 1.$$

Thus, the unexpected feature is the enhanced probability $1/(p-1)$ of divisibility by p . Once $p \mid n_k$, the distribution of the higher valuation agrees with that of a random p -adic integer conditioned to be divisible by p . Experimental results for several primes p and valuations r are recorded in Table 7.1.

To quantify how well our data matches with the proposed random model, we compute χ^2 . Concretely, for each prime p , we group the values $v_p(n_k)$ into finitely many bins. For instance, for $p = 5$, we use the seven bins $v_5(n_k) = r$ for $0 \leq r \leq 5$ and $v_5(n_k) \geq 6$. If O_i is the observed number of values in bin i , and E_i is the number predicted by the random model, then

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$

Thus a large value of χ^2 indicates a large deviation from the proposed random model. Furthermore, it is natural to consider the normalized statistic χ^2/ν , where

$$\nu = \#\{\text{Bins used in the } \chi^2\text{-test}\} - 1.$$

In Table 3.1a, we compare the positive p -valuations of our observed data with the random model $\mathbb{P}(v_p(n_k) = r) = p^{-r}$, for $r \geq 1$ (Throughout this section, $\mathbb{P}$ denotes the empirical proportion obtained by sampling k uniformly from the indicated set of parameters.) For each prime $p \geq 5$, the table gives the resulting chi-square statistic, the number of degrees of freedom, and the normalized statistic χ^2/ν . It can be seen that the prime $p = 5$ shows a strong deviation from this random model. The exact counts are recorded in Table 3.1b. The prime $p = 37$ also shows a noticeable discrepancy, although it is substantially smaller.

p	χ^2	ν	χ^2/ν
5	76.84	6	12.81
7	1.56	5	0.31
11	8.32	4	2.08
13	6.42	4	1.60
17	1.49	3	0.50
19	4.10	3	1.37
23	3.37	3	1.12
29	3.41	3	1.14
31	7.33	3	2.44
37	14.27	3	4.76
41	2.60	2	1.30
43	2.41	2	1.20
47	1.38	2	0.69

(A) Comparison of $v_p(n_k)$ to $\mathbb{P}(v_p = r) = p^{-r}$, primes $5 \leq p \leq 47$.

$v_5(n_k) = r$	count	expected	ratio
0	186,238	187,499.3	0.993
1	50,398	49,999.8	1.008
2	10,651	10,000.0	1.065
3	2,141	2,000.0	1.071
4	471	400.0	1.178
5	78	80.0	0.975
6	19	16.0	1.188
7	3	3.2	0.938

(B) Counts with $v_5(n_k) = r$ vs the expected $M \cdot 5^{-r}$, where $M := 249,999$ (the expected total at the $r = 0$ row is $M \cdot \frac{3}{4}$).

TABLE 3.1. The law $\mathbb{P}(v_p = r) = p^{-r}$ ($r \geq 1$) compared to our observed data. Left: reduced χ^2 across primes, only $p = 5$ deviates strongly. Right: exact counts $v_5(n_k) = r$.

3.3. 3-valuation observations. The 3-adic behavior of n_k does not seem to follow the same pattern as the one described in Section 3.2. We record several observations, including parity laws for the 3-valuations.

Globally higher 3-divisibility. Divisibility by 3 appears to occur considerably more frequently than the corresponding heuristic would suggest with $v_3(n_k) \geq 1$ about 73% of the time; see Table 7.2 and Figure 7.1. More specifically, the proportions of k for which $v_3(n_k) = r$ with $r \geq 2$ appear to be elevated (by at least a factor of 2) compared with the baseline prediction 3^{-r} . On the other hand, the proportion with $v_3(n_k) = 1$ appears to be slightly smaller than this baseline: for instance, our data give approximately 0.322, compared with $1/3 = 0.333 \dots$ expected from the model for $p \geq 5$. This excess is not an artifact of the congruence phenomena discussed below.

3-Divisibility among congruence classes. The 3-valuation is further structured by congruences. The proportion of k with $v_3(n_k) \geq r$ is systematically higher for $k \equiv 0 \pmod{3}$; see Table 7.3. Refining modulo 27, we find a stronger statement: for $k \equiv 0, \pm 5 \pmod{27}$ one has $v_3(n_k) \geq 1$ for all but a few exceptional k in the dataset.

Bad primes of E_k and 3-divisibility. Another observation is that there seems to be a correlation between the parity of the cardinality of the set

$$S_k := \{\text{bad primes of the form } p \equiv 1 \pmod{3} \text{ for the curve } E_k\}$$

and the 3-valuation of n_k , see Table 3.2. The table suggests that it is more likely for n_k to be divisible by 3 when the parity of $\#S_k$ is even: We have $\mathbb{P}(v_3(n_k) \geq 1 \mid \#S_k \text{ is even}) = 0.79$, compared to $\mathbb{P}(v_3(n_k) \geq 1 \mid \#S_k \text{ is odd}) = 0.67$. If we compare this to the analogous set

$$T_k := \{\text{bad primes of the form } p \equiv 2 \pmod{3} \text{ for the curve } E_k\},$$

this effect disappears. We have $\mathbb{P}(v_3(n_k) \geq 1 \mid \#T_k \text{ is even}) = 0.73$, compared to $\mathbb{P}(v_3(n_k) \geq 1 \mid \#T_k \text{ is odd}) = 0.73$. This phenomenon seems to be a consequence of a sharper law on a subfamily, which we will now describe.

	$v_3(n_k) \leq 0$	$v_3(n_k) \geq 1$	total
$\#S_k$ even	26,170 (20.9%)	98,896 (79.1%)	125,066
$\#S_k$ odd	40,811 (32.7%)	84,122 (67.3%)	124,933
total	66,981 (26.8%)	183,018 (73.2%)	249,999

	$v_3(n_k) \leq 0$	$v_3(n_k) \geq 1$	total
$\#T_k$ even	33,406 (26.7%)	91,600 (73.3%)	125,006
$\#T_k$ odd	33,575 (26.9%)	91,418 (73.1%)	124,993
total	66,981 (26.8%)	183,018 (73.2%)	249,999

TABLE 3.2. Contingency tables comparing the divisibility of n_k by 3 with the parities of $\#S_k$ and $\#T_k$. Percentages are according to the parity of $\#S_k$ and $\#T_k$.

3.3.1. *A parity law on a subfamily.* On the set

$$(3.5) \quad A := \{k : 3 \nmid k, \text{ and } k(k^2 - 16) \text{ squarefree}\},$$

the above-mentioned correlations sharpen into two laws, valid (with the exception of $k = 1$):

$$(3.6) \quad v_3(n_k) = 0 \implies \#S_k \text{ is odd}, \quad v_3(n_k) = 1 \implies \#S_k \text{ is even}.$$

see Tables 3.3 and 3.4. Here it can be seen that a similar law does not seem to appear for $v_3(n_k) \geq 2$.

$v_3(n_k)$	count	$\#S_k$ even	$\#S_k$ odd	$\mathbb{P}(\#S_k \text{ odd})$
0	13,080	1	13,079	1.000
1	13,233	13,233	0	0.000
2	8,950	4,438	4,512	0.504
3	4,161	2,055	2,106	0.506
4	1,560	780	780	0.500
5	552	266	286	0.518
6	194	95	99	0.510
7	64	31	33	0.516
8	22	10	12	0.545
9	9	5	4	0.444
10	3	1	2	0.667
11	1	0	1	1.000
12	1	1	0	0.000

TABLE 3.3. Parity of $\#S_k$ as a function of $v_3(n_k)$, for $k \in A$. The law $v_3(n_k) = 0 \implies \#S_k \text{ odd}$ and $v_3(n_k) = 1 \implies \#S_k \text{ even}$ holds with a single exception; for $v_3(n_k) \geq 2$ the parity is unconstrained.

	$v_3(n_k) = 0$	$v_3(n_k) \geq 1$	total
$\#S_k$ even	1	20,915	20,916
$\#S_k$ odd	13,079	7,835	20,914
total	13,080	28,750	41,830

TABLE 3.4. Contingency of parity($\#S_k$) against $v_3(n_k) \geq 1$ for $k \in A$.

It is apparent that law (3.6) is a special case of a more general law that holds on a larger set. Define

$$\alpha(k) := \#S_k + v_3(k) + \mathbb{1}_{8|k}.$$

The two laws are

$$(3.7a) \quad (\text{odd}) \quad v_3(n_k) = 0 \implies \alpha(k) \text{ is odd},$$

$$(3.7b) \quad (\text{even}) \quad v_3(n_k) = 1 \implies \alpha(k) \text{ is even}.$$

On A we have $k^2 - 16$ is squarefree, $v_3(k) = 0$ and $8 \nmid k$, so (3.7) reduces to (3.6). We now investigate several enlargements of the set A for which the laws (3.7) hold, see Table 3.5 and Figure 3.2. The results in Table 3.5 suggest that, among the sets considered there, the largest set on which the odd law (3.7a) holds is

$$C' := \{p^2 \nmid k^2 - 16 \text{ for all } p \equiv 3 \pmod{4}\} \setminus \{k \equiv 4 \pmod{8}\}.$$

On this set, however, the even law (3.7b) has 1,746 exceptions. Of these, 1,597 are divisible by 3^3 , and hence satisfy $v_3(n_k) \geq 1$ by the earlier remarks. For the remaining 149, the integer $k(k^2 - 16)$ is divisible by one of $\{7^3, 13^3, 19^3, 37^3, 61^3\}$, all cubes of primes congruent to 1 mod 3. This suggests that both laws (3.7) may hold on the smaller set

$$C'' = \{k \in C' : p^3 \mid k(k^2 - 16) \Rightarrow p \equiv 2 \pmod{3}\}.$$

Finally, it is worth noting that the k such that $k \equiv 0 \pmod{27}$ or $k \equiv \pm 5 \pmod{27}$ lie outside the set C'' : the former because $3^3 \mid k$ and the latter because $9 \mid k^2 - 16$ obstructs squarefreeness at the prime $3 \equiv 3 \pmod{4}$. So their elevated 3-valuation is not explained by the law on C'' .

3.4. 2-adic observations. Whenever n_k is an integer it appears to be even, the only exceptions being $k = 1, 2$ (where $n_k = 1$). More is true: among integer values of n_k , the inequality $v_2(n_k) \geq 2$ holds for all k in the dataset outside the finite set

$$k \in \{1, 2, 6, 7, 9, 20, 23, 28, 36, 48, 96, 112, 432\}.$$

The distribution of $v_2(n_k)$ is recorded in Table 7.4 and Figure 7.2.

Dependence on congruence classes. We next consider, for a fixed odd prime q , the proportion of k in a residue class $k \equiv a \pmod{q}$ with $v_2(n_k) \geq r$. For small r this proportion is close to 1 in every class, so the structure only becomes visible for larger r . Empirically, it is largest when $a \equiv \pm 4 \pmod{q}$ and second largest when $a \equiv 0 \pmod{q}$; see for example Table 7.5, where the frequency of $v_2(n_k) \geq r$ is highest for $k \equiv \pm 4 \pmod{7}$, closely followed by $k \equiv 0 \pmod{7}$. These are exactly the classes of bad reduction: since $\Delta_k = 2^{-24}k^{14}(k^2 - 16)$, we have $\Delta_k \equiv 0 \pmod{q}$ if and only if $k \equiv 0, \pm 4 \pmod{q}$.

Empirical model. From the earlier observations, the primes dividing $k^2 - 16$ (the case $k \equiv \pm 4 \pmod{q}$) have more influence than those dividing k (the case $k \equiv 0 \pmod{q}$), and empirically the appropriate

set	definition	density	(3.7a) exc.	(3.7b) exc.
A	$3 \nmid k, k(k^2 - 16)$ squarefree	0.167	0	0
A'	$3 \nmid k, k^2 - 16$ squarefree	0.184	0	46
A''	$k^2 - 16$ squarefree	0.323	0	913
B	odd part of $k^2 - 16$ squarefree	0.645	3,076	4,753
B'	$B \setminus \{k \equiv 4 \pmod{8}\}$	0.565	0	1,538
B''	$B' \cap \{p^3 \mid k \Rightarrow p \equiv 2 \pmod{3}\}$	0.536	0	0
C	$p^2 \nmid k^2 - 16$ for all $p \equiv 3 \pmod{4}$	0.721	3,423	5,330
C'	$C \setminus \{k \equiv 4 \pmod{8}\}$	0.631	0	1,746
C''	$C' \cap \{p^3 \mid k(k^2 - 16) \Rightarrow p \equiv 2 \pmod{3}\}$	0.598	0	0
	all values k	1.000	11,342	17,588

TABLE 3.5. The parity laws (3.7) on different enlargements of A , computed over the range $5 \leq k \leq 250,000$. Here “density” is the proportion of integers k that lie in the set, and “exc.” is the number of exceptions of the corresponding law.

weights are 2 and 1 respectively. We therefore assign to each odd prime q the weight

$$W_q(k) := \begin{cases} 0 & q \nmid k(k^2 - 16), \\ 1 & q \mid k, \\ 2 & q \mid k^2 - 16, \end{cases} \in \{0, 1, 2\},$$

well defined since an odd prime divides at most one of $k, k - 4, k + 4$. Writing $\omega_{\text{odd}}(n)$ for the number of distinct odd prime divisors of n , we have

$$\sum_{q \text{ odd}} W_q(k) = \omega_{\text{odd}}(k) + 2\omega_{\text{odd}}(k^2 - 16) = \omega_{\text{odd}}(k) + 2(\omega_{\text{odd}}(k - 4) + \omega_{\text{odd}}(k + 4)),$$

and find a moderately strong positive correlation (Pearson $r = 0.69$) between this quantity and $v_2(n_k)$. Since this sum involves only the odd primes, it appears to be natural to correct it by a term $c(k)$ that incorporates the local behavior of E_k at the prime 2. Let

$$(3.8) \quad c(k) = \begin{cases} 4, & E_k \text{ semistable at } 2 \text{ } (v_2(N_k) \in \{0, 1\}), \text{ i.e. } k \text{ odd or } 16 \mid k, \\ 2, & E_k \text{ additive at } 2 \text{ } (v_2(N_k) \in \{3, 4\}), \text{ otherwise.} \end{cases}$$

We then set

$$(3.9) \quad \hat{v}(k) := \sum_{q \text{ odd}} W_q(k) - c(k) = \omega_{\text{odd}}(k) + 2\omega_{\text{odd}}(k^2 - 16) - c(k).$$

Then $v_2(n_k)$ and $\hat{v}(k)$ have Pearson correlation $r = 0.77$ (Figure 3.3), improving on the value $r = 0.69$ obtained without the correction $c(k)$. In fact, $\hat{v}(k)$ seems to underestimate $v_2(n_k)$ for all but finitely many k . This leads us to the following conjecture.

Conjecture 3.2. *For all $k \neq 4, 8$, we have*

$$\hat{v}(k) \leq v_2(n_k).$$

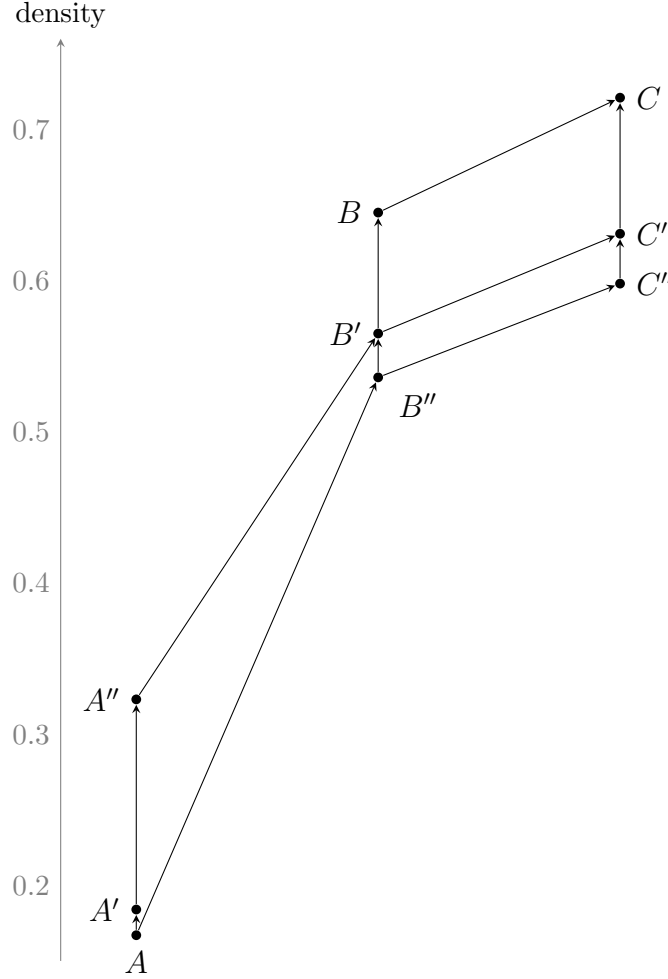


FIGURE 3.2. The sets of Table 3.5, placed at height equal to their density. An arrow points from a set to a larger set containing it; since a subset is never denser than its superset, every arrow points upward.

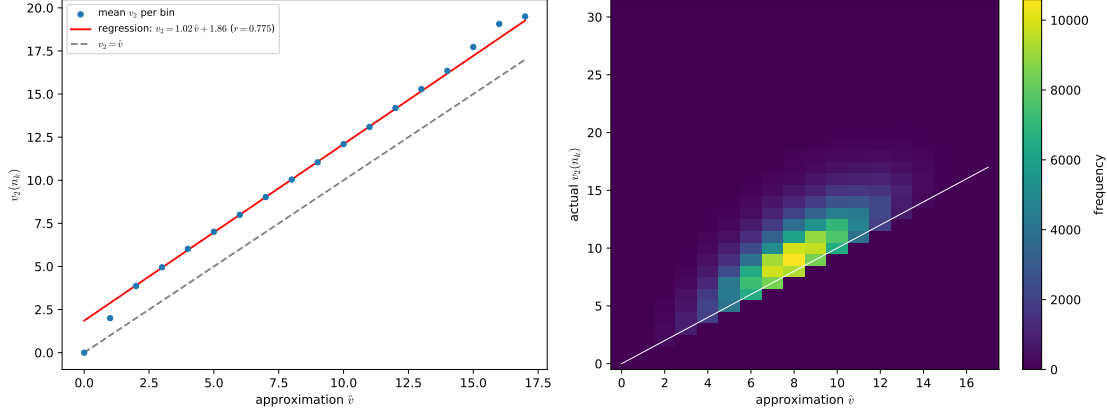
Moreover, the excess $G(k) := v_2(n_k) - \hat{v}(k)$ has mean 2.05 and the data suggest that $G(k) = 0$ may occur infinitely often, making the conjecture sharp in that sense. This excess seems to spike whenever $v_2(k) = 3$, and then drops down below 2 when $v_2(k) \geq 4$ (see Table 3.6a). This spike comes from $k \equiv \pm 8 \pmod{64}$, where the mean excess is 3.02 (see Table 3.6b). In fact it grows further as $k \equiv \pm 8 \pmod{2^m}$ for increasing $m \geq 6$, plateauing at an excess of 3.5 (see Table 3.6c). A reasonable fit for the mean excess $G(k)$ seems to be

$$(3.10) \quad \mathbb{E}[G(k)] \approx 2 + s(k), \quad s(k) := \frac{3}{2} \max \left\{ 1 - 2^{5 - \max\{v_2(k-8), v_2(k+8)\}}, 0 \right\}.$$

Here $s(k)$ vanishes when k is not 2-adically close to ± 8 and approaches 1.5 as k approaches ± 8 2-adically.

Combining these observations, our empirical model is

$$(3.11) \quad v_2(n_k) = \hat{v}(k) + s(k) + X_k,$$

FIGURE 3.3. $v_2(n_k)$ vs $\hat{v}(k)$ for k with $v_2(n_k) \geq 0$

class	count	mean $v_2(n_k)$	mean $G(k)$
k odd	125,000	9.56	+2.05
$v_2(k) = 1$	62,500	11.28	+2.04
$v_2(k) = 2$	31,249	10.39	+2.00
$v_2(k) = 3$	15,625	11.29	+2.41
$v_2(k) = 4$	7,813	8.67	+1.86
$v_2(k) = 5$	3,906	8.65	+1.91
$v_2(k) = 6$	1,953	8.55	+1.90
$v_2(k) = 7$	977	8.52	+1.93
$v_2(k) \geq 8$	976	8.33	+1.95
all	249,999	10.15	+2.05

(A) $v_2(n_k)$ by the 2-adic class of k .

$k \bmod 64$	count	mean $v_2(n_k)$	mean $G(k)$
8	3,907	11.96	+3.09
24	3,906	10.69	+1.81
40	3,906	10.70	+1.82
56	3,906	11.83	+2.94
all $v_2(k) = 3$	15,625	11.29	+2.41

(B) The class $v_2(k) = 3$, refined by $k \bmod 64$.

M	count	mean $v_2(n_k)$	mean $G(k)$	$2 + s(k)$
5	7,812	10.69	+1.81	2.00
6	3,906	11.53	+2.65	2.75
7	1,954	12.19	+3.31	3.13
8	976	12.30	+3.43	3.31
9	488	12.39	+3.50	3.41
≥ 10	489	12.37	+3.47	$\rightarrow 3.5$

(C) Excess by exact 2-adic depth $M := \max\{v_2(k-8), v_2(k+8)\}$, compared with the model mean $2 + s(k)$.TABLE 3.6. 2-adic valuation $v_2(n_k)$, with excess $G(k) = v_2(n_k) - \hat{v}(k)$. Left: grouped by the 2-adic class of k . Right: the exceptional class $v_2(k) = 3$, split further modulo 64. Bottom: the exceptional classes by 2-adic depth $M = \max\{v_2(k-8), v_2(k+8)\}$.

where the residual term X_k appears to have a distribution that is largely independent of k . More precisely, its empirical distribution is well approximated by

$$X_k \approx \text{NB}\left(r = 3, p = \frac{3}{5}\right),$$

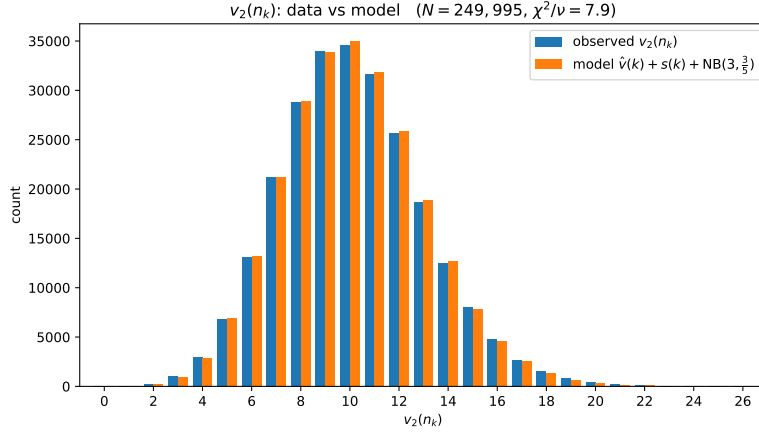


FIGURE 3.4. The observed distribution of $v_2(n_k)$ for values with $v_2(n_k) \geq 0$, compared with the prediction $\hat{v}(k) + s(k) + \text{NB}(3, \frac{3}{5})$.

where NB denotes the *negative binomial distribution*. Accordingly, in this probabilistic model, we take

$$\mathbb{P}(X_k = j) = \binom{j+2}{2} \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^j, \quad j \geq 0.$$

This distribution has mean 2 and mode 1. A comparison between this statistical model and our data is shown in Figure 3.4.

Since the mode of this distribution is 1, the best deterministic predictor we found for $v_2(n_k)$ is

$$(3.12) \quad \hat{P}(k) := \text{round}(\hat{v}(k) + s(k) + 1).$$

This predictor has Pearson correlation coefficient 0.78 with $v_2(n_k)$, only slightly higher than the correlation obtained using $\hat{v}(k)$ alone. Moreover, the prediction $\hat{P}(k)$ is correct 24.3% of the time.

Remark 3.3. Since $\hat{v}(k) + c(k) = \sum_{q \text{ odd}} W_q(k)$ is expressed as a sum of local contributions over the odd primes, and the conditions $q \mid k$ and $q \mid (k^2 - 16)$ hold on 1 and 2 residue classes modulo q (with densities $1/q$ and $2/q$), each odd prime contributes $1 \cdot \frac{1}{q} + 2 \cdot \frac{2}{q} = \frac{5}{q}$ to $\hat{v}$ on average. By Mertens' theorem, and since $c(k)$ is bounded,

$$\frac{1}{x} \sum_{4 \leq k \leq x} \hat{v}(k) = \sum_{q \leq x} \frac{5}{q} + O(1) = 5 \log \log(x) + O(1)$$

as $x \rightarrow \infty$. So, assuming Conjecture 3.2, the average order of $v_2(n_k)$ is at least $5 \log \log(x)$. Tables 7.6 and 7.7 list some of the maximisers of $W := \omega_{\text{odd}}(k) + 2\omega_{\text{odd}}(k^2 - 16)$ in the window $2.5 \times 10^5 < k < 2 \times 10^6$. In every case $v_2(n_k) \geq \hat{v}(k)$, with equality at $k = 1540136$, $k = 1812794$, and $k = 1984814$. We also remark the large value $v_2(n_k) = 27$ occurring at $k = 1069814$ (though the largest value we observe is $v_2(n_k) = 31$, at the much smaller $k = 36326$).

3.5. The regime $k = \sqrt{m}$. For $m < 0$ the parameter $k = \sqrt{m}$ is imaginary and (2.1) does not apply directly; instead we use the functional equation [LR07, Theorem 2.2] (taking $k = (\sqrt{-m} - \sqrt{16 - m})/4$ in their notation). Using the notation $\mu(k) := \mu_k$, giving

$$\mu(\sqrt{m}) = \mu\left(\frac{64}{(\sqrt{-m} - \sqrt{16 - m})^2}\right) - \mu(\sqrt{16 - m}), \quad m < 0,$$

allowing us to compute r_k also for $m < 0$.

We have computed these r_k for $|m| \leq 30,000$. We briefly compare the observations with integer values of k . First of all, for

$$m \notin \{-768, -16, -9, -2, -1, 9, 18, 25, 27, 64, 144, 256, 1024\},$$

r_k is the reciprocal of an integer. For $p \geq 5$, the law $\mathbb{P}(v_p(n_k) = r) \approx p^{-r}$ holds unchanged. For $p = 3$, the parity law of Section 3.3.1 breaks down completely (it holds for only about half of the relevant k); of the accompanying congruences, only $m \equiv 25 \pmod{27}$ survives, that is, the image of $k \equiv \pm 5$. This forces once again $v_3(n_k) \geq 1$.

For $p = 2$, a variant of the lower bound in Conjecture 3.2 holds. The term $\omega_{\text{odd}}(k)$ appears to split into two different counts on m

$$\omega_{\text{odd}}^0(m) + 2\omega_{\text{odd}}^1(m),$$

where $\omega_{\text{odd}}^0(m)$ is the number of odd primes p dividing m with $v_p(m)$ even, and $\omega_{\text{odd}}^1(m)$ the number of odd primes with $v_p(m)$ odd. Together with a sign-dependent shift that makes the bound sharp, the generalization of $\hat{v}(k)$ becomes

$$\hat{v}(m) = \omega_{\text{odd}}^0(m) + 2\omega_{\text{odd}}^1(m) + 2\omega_{\text{odd}}(m - 16) - c(m) + \varepsilon(m),$$

where

$$c(m) = \begin{cases} 4, & v_2(m) = 0 \text{ or } v_2(m) \geq 8, \\ 2, & 1 \leq v_2(m) \leq 7, \end{cases} \quad \varepsilon(m) = \begin{cases} 0, & |m| \text{ a perfect square,} \\ 1, & m < 0 \text{ nonsquare,} \\ 2, & m > 0 \text{ nonsquare.} \end{cases}$$

Then $\hat{v}(m) \leq v_2(n_k)$ appears to hold for all m with integral n_k , the sole exception being $m = 12$. Equality is attained for about 17% of these m ; the residual $v_2(n_k) - \hat{v}(m)$ is then non-negative, of the same negative-binomial flavor as the excess X_k in the integer case. We further remark that all of the behavior described above also holds when m is restricted to a single sign.

3.6. A conditional proof of Conjecture 3.2 for special families of k . We now use the work of Dummigan, Golyshev, de Jeu and Kerr [DGdJK26] to verify Conjecture 3.2 conditionally on the Bloch–Kato conjecture and the additional hypotheses appearing in their work. Let

$$\tilde{E}_u: y^2 = x(x+1)(x+u^2).$$

The curve $\tilde{E}_u$ is 2-isogenous to E_{4u} ; hence, when $k = 4u$, the two curves have the same L -function and conductor.

Proposition 3.4. *Let $k = 4u$, where $u \equiv 4 \pmod{8}$ and $u(u^2 - 1)$ is squarefree at the odd primes; assume in addition that $u + 1$ has a prime divisor $\equiv 3 \pmod{4}$. Assume moreover, as in [DGdJK26, Thm. 3], that $K_2^T(\tilde{E}_u)_{\text{int}} \otimes_{\mathbb{Z}} \mathbb{Q}$ is one-dimensional, that r_k is a rational number and that $H_f^1(\mathbb{Q}, \tilde{E}_u[2^\infty])(-1)$ is finite; and assume the Bloch–Kato conjecture for $\tilde{E}_u$.*

Then Conjecture 3.2, i.e. $\hat{v}(k) \leq v_2(n_k)$, is true for these k .

Proof. By the hypotheses on u , the hypothesis of [DGdJK26, Theorem 1] are satisfied and $m_u = 1$, so that $\iota_{u,2} = 1$. In the notation of [DGdJK26], we have $F(u) = \mu_k$ and

$$q_u = \frac{L(\tilde{E}_u, 2)}{(2\pi i)^2 F(u)}.$$

By the functional equation and the fact that $\tilde{E}_u$ is isogenous to E_k , one has

$$n_k = -\omega_k N_k q_u.$$

Under the present hypotheses, the conductor is

$$N_k = \frac{u(u^2 - 1)}{4},$$

which is odd. Consequently,

$$v_2(n_k) = v_2(q_u).$$

On the one hand, since $\iota_{u,2} = 1$, [DGdJK26, Equation (1.3)] reads

$$(3.13) \quad v_2(n_k) + 2 = 2\omega_1(u) + \omega_3(u) + \omega(u^2 - 1) + s_u,$$

where

$$s_u = v_2 \left(\#H_f^1(\mathbb{Q}, \tilde{E}_u[2^\infty](-1)) \right),$$

ω is the number of distinct prime factors, and ω_j the number of distinct prime factors $\equiv j \pmod{4}$. On the other hand, $k = 4u$ implies $\omega_{\text{odd}}(k) = \omega_{\text{odd}}(u) = \omega_1(u) + \omega_3(u)$ and similarly, $\omega_{\text{odd}}(k^2 - 16) = \omega(u^2 - 1)$. Since $16|k$, we have $c(k) = 4$. Hence

$$(3.14) \quad \hat{v}(k) = \omega_1(u) + \omega_3(u) + 2\omega(u^2 - 1) - 4.$$

Combining the two Equations (3.13) and (3.14) gives

$$v_2(n_k) - \hat{v}(k) = s_u - (\omega(u^2 - 1) - \omega_1(u) - 2).$$

Thus, in order to prove our conjecture for these k , it suffices to prove the inequality

$$s_u \geq \omega(u^2 - 1) - \omega_1(u) - 2.$$

By elementary group theory we have that

$$s_u \geq s'_u := v_2(\#H_f^1(\mathbb{Q}, \tilde{E}_u[2^\infty](-1))[2]).$$

We now use the explicit description of this 2-torsion subgroup given in [DGdJK26, Theorem 2]. Let S be the set of prime divisors of $u^2 - 1$, and let S' be the set of prime divisors of u that are congruent to 1 mod 4. Restricting the parametrization in [DGdJK26, Theorem 2] to pairs of the form $(D, 1)$ as discussed in [DGdJK26, Remark 7.9], we obtain a subspace consisting of the squarefree divisors D of $u^2 - 1$ such that

$$D \text{ is a square modulo every prime in } S' \quad \text{and} \quad D \equiv 1 \pmod{8}.$$

Writing $D = \prod_{p \in S} p^{x_p}$ with $x_p \in \mathbb{F}_2$, the above two conditions give a linear system in the $\omega(u^2 - 1)$ variables x_p . Indeed, if $\left(\frac{p}{q}\right) = (-1)^{\alpha_{p,q}}$ for $\alpha_{p,q} \in \mathbb{F}_2$ for each prime q dividing u such that $q \equiv 1 \pmod{4}$, we get a system of $\omega_1(u)$ equations $\sum_p x_p \alpha_{p,q} = 0$ in $\omega(u^2 - 1)$ variables for the first condition. The second condition can be expressed as $\left(\frac{-1}{D}\right) = \left(\frac{2}{D}\right) = 1$, and this gives at most two extra equations, $\sum_p x_p \beta_{p,-1} = \sum_p x_p \beta_{p,2} = 0$, where $\left(\frac{-1}{p}\right) = (-1)^{\beta_{p,-1}}$ and $\left(\frac{2}{p}\right) = (-1)^{\beta_{p,2}}$. Thus, the system has $\omega(u^2 - 1)$ variables and at most $\omega_1(u) + 2$ equations. Its solution space therefore has dimension at least $\omega(u^2 - 1) - \omega_1(u) - 2$. Consequently,

$$s_u \geq \omega(u^2 - 1) - \omega_1(u) - 2,$$

and we recover that statement of Conjecture 3.2 under these particular conditions. $\square$

Remark 3.5 (The Bloch–Kato formula at an arbitrary prime). The formulation in [DGdJK26, §12] is not restricted to the prime 2. Whenever the hypotheses in Assumption 12.1 of [DGdJK26] hold, the Bloch–Kato conjecture predicts

$$\begin{aligned} v_p(n_k) = v_p(N_k) &+ \sum_{q \leq \infty} v_p \left(\text{Tam}_{q,\omega}^0(H_{\text{ét}}^1(\tilde{E}_{u,\overline{\mathbb{Q}}}, \mathbb{Z}_p)(2)) \right) \\ &+ v_p \left(\#H_f^1(\mathbb{Q}, \tilde{E}_u[p^\infty](-1)) \right) \\ &- v_p \left(\#H^0(\mathbb{Q}, \tilde{E}_u[p^\infty](1)) \right) - v_p \left(\#H^0(\mathbb{Q}, \tilde{E}_u[p^\infty](-1)) \right) \\ &- v_p(\iota_{u,p}), \end{aligned}$$

where $k = 4u$ and $\iota_{u,p}$ is the regulator index defined in [DGdJK26, §12]. Thus, conjecturally, the valuation of n_k decomposes into conductor and local Tamagawa contributions, a Tate-twisted Bloch–Kato Selmer contribution, global invariant terms, and a regulator index.

At present, however, this does not give a practically computable formula at a general odd prime: neither the full p -primary Bloch–Kato Selmer group nor the regulator index $\iota_{u,p}$ is known in general. For example, when $u = 4$, the identity $n_{16} = 1/11$ is interpreted in [DGdJK26] as predicting that

$$11 \mid \iota_{4,11}.$$

4. THE MODEL

We train a transformer-based model [VSP⁺17] using AXOLVER, a system based on Int2Int [Cha25], on our dataset. The transformer-based model has 4 encoder layers and 4 decoder layers, 8 attention heads and an embedding dimension of 256 on batches of 32 examples, using the Adam optimizer [KB14] with a constant learning rate of 10^{-4} . All other parameters were left at their default values unless otherwise indicated.

Our input is an integer k , encoded as a sequence of digits. The target output is a rational number r_k , written as a pair of integers, each again encoded as a sequence of digits. Thus, the task is to learn

$$k \mapsto r_k$$

from our dataset.

We randomly split our set into *training*, *validation* and *test sets*, of sizes 243,999, 3,000 and 3,000, respectively. The training set is used to train the neural network, while the test set is used to monitor how well the model learns over the course of training. Since the hyperparameter tuning did not yield any improvements, we did not use the validation set in our analysis. We measure the accuracy on this set by the *greedy accuracy*, which is the percentage of times the model guesses correctly. We evaluate the model performance at the end of every epoch which we set to be equal to 1,000 steps.

We attempt to learn r_k in three types of tasks: learning r_k directly, learning the size of $|r_k|$, and learning the p -adic valuations $v_p(n_k)$. The tasks involving the learning of r_k and of the size of $|r_k|$ are performed on base 10 (as a default), but the tasks involving the learning of $v_p(n_k)$ are performed on base p , since this allows the model to detect multiplicity by p and the value of v_p more easily than in other bases. For the base 10 tasks, we use `max_len` = 256 and `max_output_len` = 64. For the valuation tasks, which are performed on bases 2, 3, 5, and 7, we use `max_len` = `max_output_len` = 512 as the tokenized data sets are too large to be handled with smaller parameters.

We run the experiments with different features (computed with PARI/GP [PAR25]) consisting of various arithmetic invariants of the elliptic curve E_k . These are as follows.

- ω_k , the root number, which is the sign of the functional equation of $L(E_k, s)$. As explained in the introduction, this value is expected to codify the sign of r_k .
- N_k and $\text{factor}(N_k)$, the conductor of E_k and its factorization. Our goal is to measure whether information on prime reduction can help the model determine arithmetic properties of r_k . We are also interested in seeing if the model can learn the size $|r_k|$ from N_k , in view of the estimate (3.4).
- Δ_k and $\text{factor}(\Delta_k)$, the discriminant of E_k given by the model (1.1), and its factorization. Its formula is $\Delta_k = 2^{-24}k^{14}(k^2 - 16)$. Once again, we are interested in studying the connection of prime reduction and the arithmetic properties of r_k . Moreover, as equations (3.9) and (3.11) indicate, the conjectural estimate for $v_2(n_k)$ seems to involve a count on the primes of bad reduction, with odd primes counted with different weight according to the type of reduction, which is better encoded by the exponents codified in the factorization of the discriminant.
- a_p , the trace of Frobenius for $p = 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37$, given as a vector. We wish to observe if giving the first traces of Frobenius can help the model learn r_k , given that they codify the first coefficients of the L -function.
- The rank of E_k . This was computed with the `ellrank` command in PARI/GP [PAR25] and was therefore represented as a vector $[\text{rank}_{\min}, \text{rank}_{\max}]$. While the rank is supposed to codify information about $L(E_k, s)$ at $s = 1$ and not $s = 2$ (via the Birch and Swinnerton-Dyer conjecture), this is still a natural quantity to consider due to its importance.
- c_2 , the Tamagawa number at $p = 2$. We consider this quantity in the learning of $v_2(n_k)$, due to its role in codifying information on the reduction of E_k at 2, which is very different from the situation of the other primes.

5. MACHINE LEARNING RESULTS

In this section we analyze the output of our neural network, and compare it to the results of the previous section 3.

5.1. The size of $|n_k|$. The initial experiments intended for the model to learn r_k given various features discussed in Section 4 (all written in base 10). However, this approach proved to be too ambitious, as the experiments yielded validation accuracy essentially 0. For these experiments, we noticed that the model only seems to learn the size of r_k (or equivalently, the size of n_k). In the discussion that follows, we will refer to *learning the size of $|n_k|$* to the experiment where the model output s_k is considered correct if s_k has the same sign as r_k and if $\frac{|r_k|}{2} < |s_k| < |2r_k|$, that is, the prediction of the *factor-two accuracy* for r_k .

The first observation is that including the root number ω_k as a feature doubles the accuracy, see Figure 5.1, which suggests that the model does not learn anything about the sign of r_k unless it is provided with said sign as a feature. Because of this, we conducted the experiments almost always including ω_k as a feature.

In the context described above, we found that the model learned best when given the conductor N_k as a feature. In this case, as well as in the cases where the model was given the factors of the conductor $\text{factor}(N_k)$ and the factors of the discriminant $\text{factor}(\Delta_k)$, the model reached 100% accuracy. The fastest regime for reaching this level was with N_k , closely followed by $\text{factor}(\Delta_k)$ and closely followed by $\text{factor}(N_k)$. These findings are recorded in Figure 5.2. Given the estimate (3.4) we are not surprised by this result, as $\frac{N_k}{\log(k)}$ essentially codifies the size of n_k , and $\log(k)$ is a relatively small quantity. Moreover, we speculate that the model recovers N_k from $\text{factor}(\Delta_k)$ and $\text{factor}(N_k)$ and is unable to recover N_k from Δ_k (due to the fact that factorization is difficult).

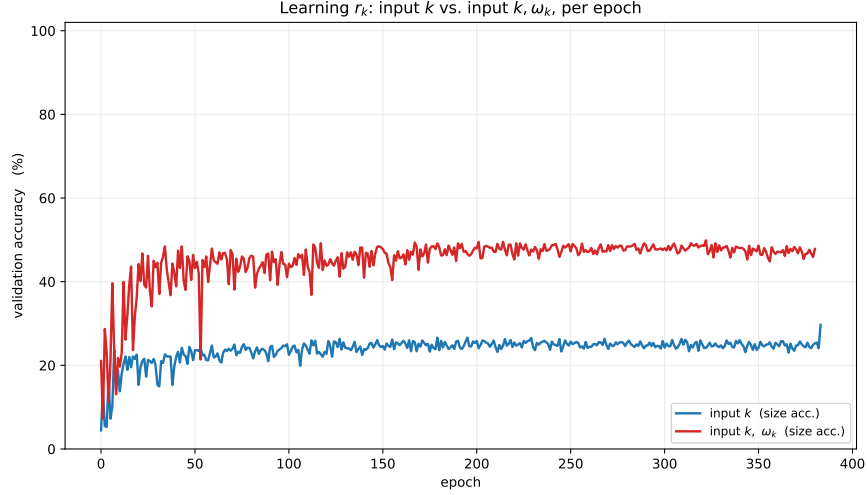


FIGURE 5.1. The model's accuracy on the size of n_k with ω_k included versus not included.

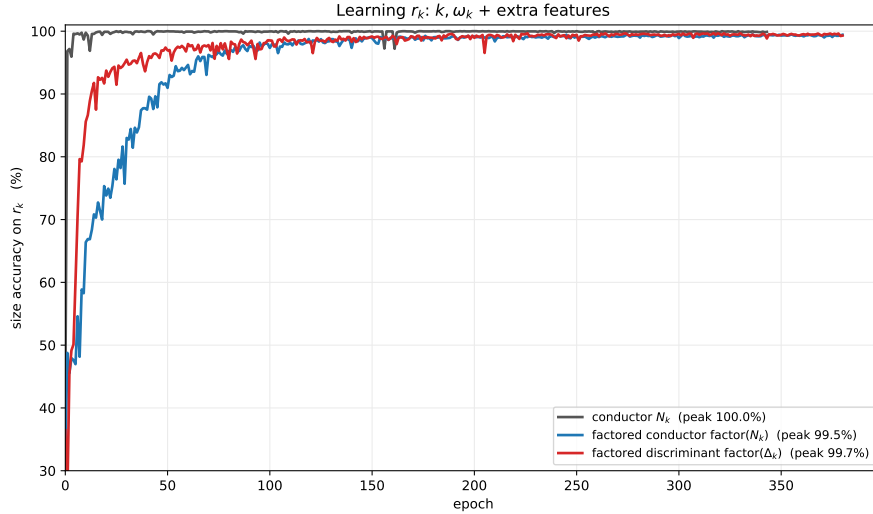


FIGURE 5.2. Learning r_k from k and ω_k in three variants: Giving N_k , $\text{factor}(N_k)$ or $\text{factor}(\Delta_k)$.

Nevertheless, we find it puzzling that $\text{factor}(\Delta_k)$ leads to faster learning than $\text{factor}(N_k)$. The other considered features, namely the rank and the a_p 's do not seem to influence the model's learning.

Finally, in Figure 5.3 we see that the values of $L(E_k, 2)$ induced by the model's predictions have a marginal distribution resembling the empirical distribution of the true values from Figure 3.1. This does not imply accurate pointwise prediction of $L(E_k, 2)$. Interestingly, the experiment given by k, ω_k and $\text{factor}(\Delta_k)$ seems to reproduce this distribution the best.

5.2. p -adic valuations of n_k for $p \geq 5$. These experiments were performed for $p = 5, 7$. They had $v_p(n_k)$ as output with k as a lone feature, and also with additional feature $\text{factor}(N_k)$, as well as additional feature $\text{factor}(\Delta_k)$ (only for $p = 5$), all written in base p . In all cases the model

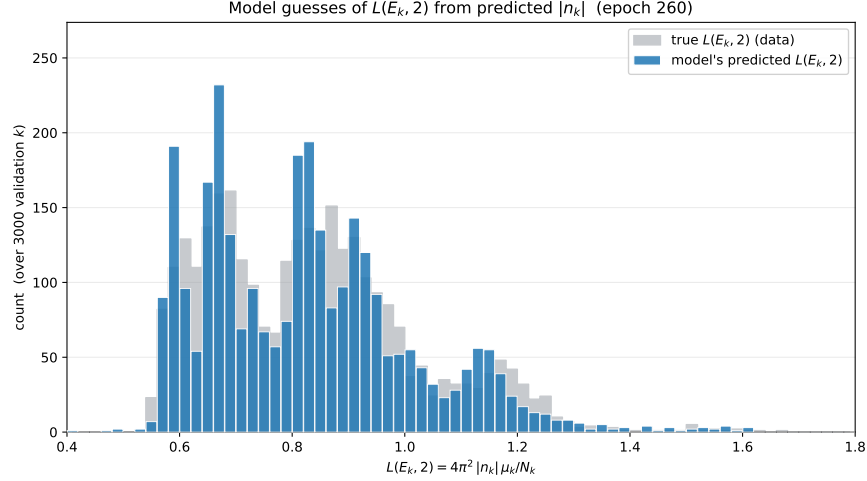


FIGURE 5.3. Distribution of the model guesses for $L(E_k, 2)$ against the true values of $L(E_k, 2)$, for k in the validation set. This is for the experiment where k, ω_k and $\text{factor}(\Delta_k)$ are given as input.

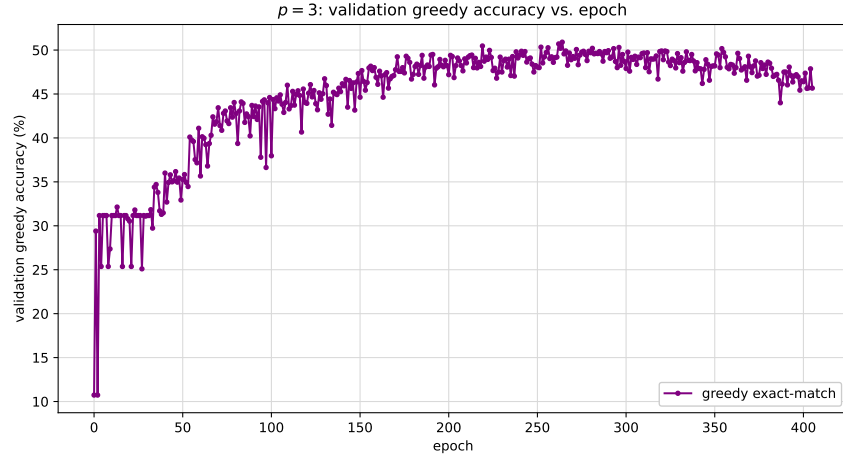
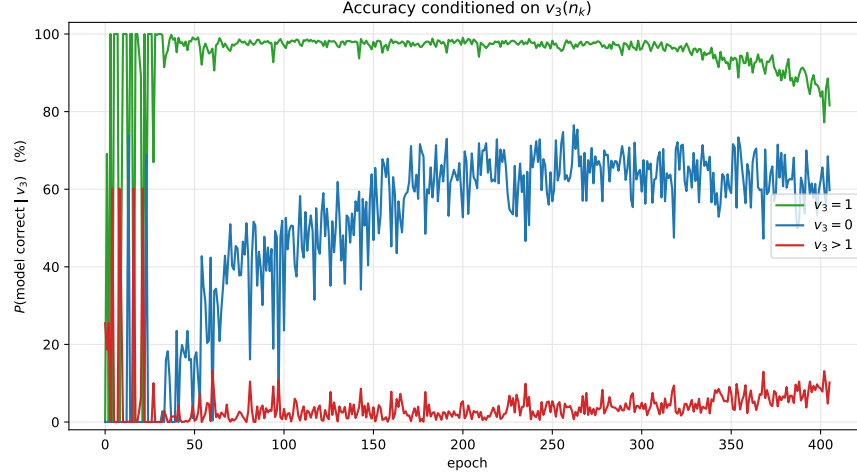
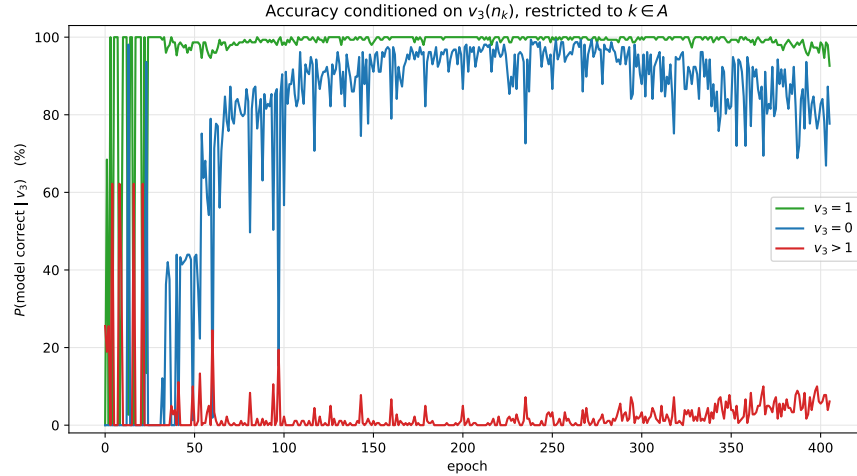


FIGURE 5.4. The validation accuracy (i.e. the percentage of correct guesses by the model) as a function of the epochs.

eventually predicts that none of the n_k are multiple of p , thus leading to a greedy accuracy of $\frac{p-2}{p-1}$. Therefore, we cannot extract any conclusions from these experiments.

5.3. 3-adic valuations of n_k . For $p = 3$, we found that the model learned best when given k and the factors of the discriminant $\text{factor}(\Delta_k)$ as features and $v_3(n_k)$ as output, all written in base 3. We recorded the validation accuracy in Figure 5.4. From this figure, we learn that the accuracy maximizes near 51% around epoch 264. Moreover, we see that the model starts overfitting from roughly epoch 300 on. In this pre-overfitting phase, the model seems to almost always guess that $v_3(n_k) \in \{0, 1, 2, 3\}$ (this happens 99.99% of the time, on average). The accuracy for n_k with high 3-valuations, $v_3(n_k) \in \{2, 3\}$ is quite low, see Figure 5.5. In this figure, we learn that validation accuracy of the model conditioning $v_3(n_k) = 1$ is nearly 100% all the time. Though this is a bit misleading: from the earliest epochs the model simply guesses $v_3(n_k) = 1$ for nearly every input,

FIGURE 5.5. The validation accuracy, conditioned on $v_3(n_k)$ FIGURE 5.6. The validation accuracy on $k \in A$, conditioned on $v_3(n_k)$.

and so trivially classifies the true $v_3(n_k) = 1$ cases correctly (this also explains the starting plateau in Figure 5.4). The surprising fact is that this near 100% accuracy on $v_3 = 1$ persists even as the model learns to recognize $v_3 = 0$ cases, and the corresponding curve climbs.

5.3.1. Learning on different congruence classes. As observed in the previous section, we have that $v_3(n_k) \geq 1$, whenever $k \equiv 0, \pm 5 \pmod{27}$. We have that the model quite quickly learns this fact: $\mathbb{P}(\text{model predicts } v_3(n_k) = 0 \mid k \equiv 0, \pm 5 \pmod{27})$ exceeds 1% only at 7 epochs in the entire run. In fact, something more concerning happens: the model predicts almost always (i.e. 99.3% of the time in epoch range 100 – 300) that $v_3(n_k) = 1$ for those k . So, even though these congruence classes look more deterministic, their accuracy is in fact the lowest among all the congruence classes (their accuracy at epoch 264 is just 41.5%, compared to the global accuracy at this epoch being 50.9%).

5.3.2. Learning on the sets from Table 3.5. Next we study how well the model learns the laws (3.7) for $k \in A$. In Figure 5.6 we measure the validation accuracy assuming $v_3(n_k) = 0, 1$ or > 1 . From this figure, we learn that the accuracy for $v_3(n_k) > 1$ is still very poor, but that the

set	#	acc.
A	306	99.3%
A'	340	97.4%
A''	566	95.1%
B	1,133	89.8%
B'	977	95.2%
B''	937	95.1%
C	1,251	88.6%
C'	1,081	93.8%
C''	1,035	93.6%

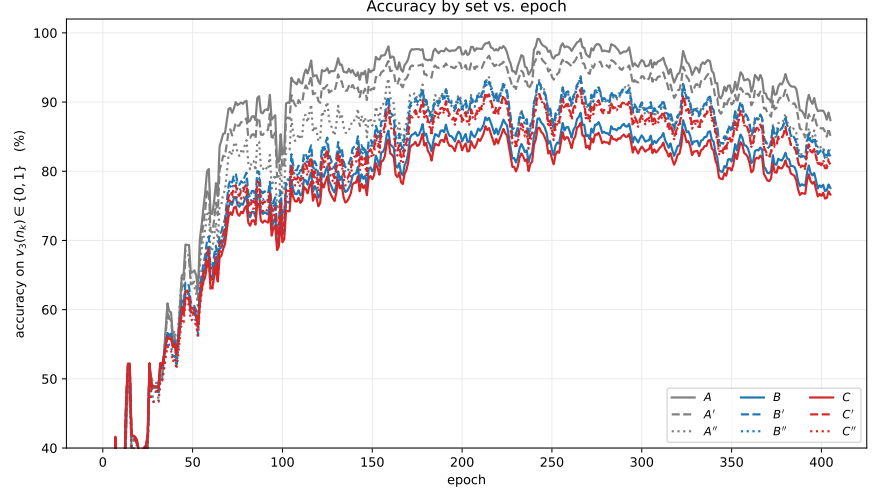


FIGURE 5.7. Accuracy on $v_3(n_k) \in \{0, 1\}$ over the sets of Table 3.5. Left: accuracy at the peak epoch 264, with n the number of validation k in each set. Right: the same accuracies as a function of the epochs.

accuracy assuming $v_3(n_k) = 0$ is 98.7%. Thus, on the set A , the model is highly successful in distinguishing $v_3(n_k) = 0$ from $v_3(n_k) = 1$. Figure 5.7 shows the model's validation accuracy, restricted to $v_3(n_k) \in \{0, 1\}$, on each of the sets in Table 3.5, both at the peak epoch and as a function of training. In this figure, we see that this conditional accuracy is the highest on the set A . Although the parity laws remain exact on some of the larger sets, these sets admit more varied factorizations of $k(k^2 - 16)$. We speculate that this makes the relevant parity information more difficult for the model to extract, leading to lower accuracy. Removing the values $k \equiv 4 \pmod{8}$, however — passing from B to B' and from C to C' — does increase this accuracy again.

Finally, we consider the complement of $C'' \cup \{k \equiv 0, \pm 5 \pmod{27}\}$, i.e. the values not covered by our largest set or the congruence classes above. As Figure 5.8 shows, the model accuracy is still increasing on the set. This leads us to suspect the model is still learning some arithmetic features, beyond the ones we already discussed.

After the experiment giving $\text{factor}(\Delta_k)$, the model learns the most if given the factors of the conductor $\text{factor}(N_k)$. The model learns faster from $\text{factor}(N_k)$ than from $\text{factor}(\Delta_k)$. Giving both $\text{factor}(N_k)$ and $\text{factor}(\Delta_k)$ together as features does not lead to a significant difference in learning compared to only giving $\text{factor}(N_k)$, but the learning is slower. The model does not exhibit learning if given N_k or Δ_k as features. We speculate that this is due to the difficulty in factorization.

5.4. 2-adic valuations of n_k . For $p = 2$, we found that the model learned best when given k and the factorization of the discriminant $\text{factor}(\Delta_k)$ as input and $v_2(n_k)$ as output, all written in base 2. The validation accuracy is recorded in Figure 5.9. We see that the model's overfitting starts happening (i.e. the validation accuracy decreases) around epochs 220 – 240. From the figure we learn that the model climbs almost immediately to an accuracy of about 23% – 25% in the first few epochs and has a *first* plateau there from roughly epoch 5 to 75. It then breaks through and starts climbing through the epochs 80 – 150. After that it settles at a *second* plateau with accuracy of roughly 36% – 38%, maximizing at epoch 210 with an accuracy 39.6%. We analyze the behavior of the model at these plateaus.

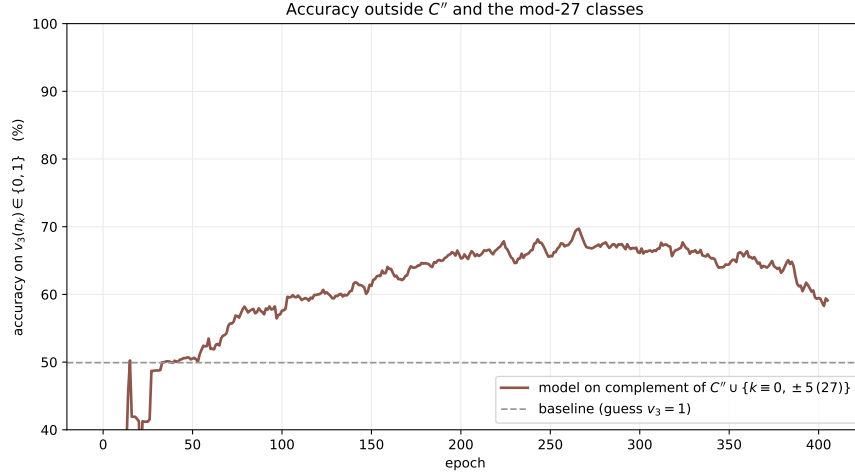


FIGURE 5.8. The accuracy assuming $v_3(n_k) \in \{0,1\}$ outside C'' and $k \equiv 0, \pm 5 \pmod{27}$.

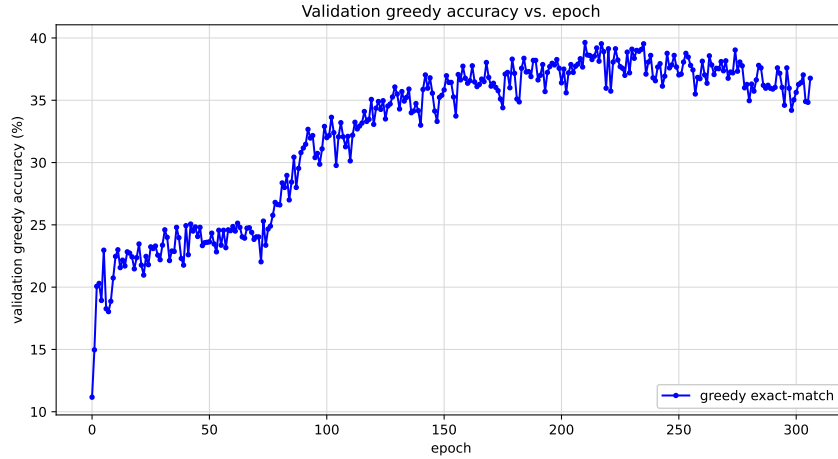


FIGURE 5.9. The validation accuracy (i.e. the percentage of correct guesses by the model) as a function of the epochs.

5.4.1. *Learning around the first plateau of Figure 5.9.* We first investigate whether the model learns our conjectural lower bound $\hat{v}(k) \leq v_2(n_k)$ from Conjecture 3.2. As Figure 5.10 shows, from epoch 8 onward the model's predictions respect this bound for almost all k in the validation set.

In particular, we compare our predictor $\hat{P}(k) = \text{round}(\hat{v}(k) + s(k) + 1)$ (given in (3.12)) with the model's prediction by computing the Pearson correlation coefficient between the two as a function of the epochs, see Figure 5.12. We see that this correlation rises quickly and reaches a maximum of $r = 0.97$ around epochs 40 – 70. As it leaves the first plateau, the correlation gradually drops to around $r \approx 0.88$ in the last epoch. Since the accuracy rises after the first plateau, but the correlation is dropping, this means that the model learns something beyond our predictor $\hat{P}(k)$. For contrast, the correlation between the true valuations $v_2(n_k)$ and our predictor over the full dataset of 249,999 values is only $r = 0.78$. This behavior is illustrated in Figure 5.11a, where we plot the confusion matrix of the model's prediction against our deterministic predictor $\hat{P}(k)$. It can

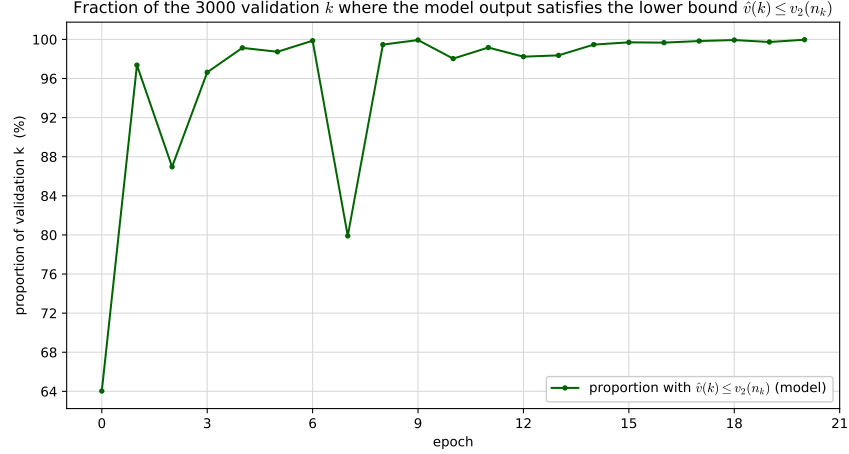


FIGURE 5.10. The proportion of k in the validation set for which the model predicts $\hat{v}(k) \leq v_2(n_k)$.

be seen that the mass of the confusion matrix lies close to the diagonal. As the number of epochs grows further, the distribution spreads out from the diagonal even before overfitting. Together with the simultaneous increase in accuracy, this suggests that the model learns additional structure beyond $\hat{P}(k)$. Figure 5.11b shows the analogous confusion matrix of the true valuation $v_2(n_k)$ against the model's prediction. Since the correlation between $\hat{P}(k)$ and the model's prediction is very close to 1 throughout the first plateau, we conclude that during this phase the model has essentially learned to approximate $\hat{P}(k)$. As guessing $\hat{P}(k)$ is correct only 24.3% of the time, and the model's accuracy on the first plateau is likewise about 23% – 25%, it appears that at this stage the model has learned nothing beyond this predictor.

5.4.2. Learning around the second plateau of Figure 5.9. For the second plateau, it becomes much less clear *what* the model actually learned. But there are several hints, for instance, the main improvement from the first to the second plateau is to detect when the lower bound in Conjecture 3.2 is sharp, i.e. when $v_2(n_k) = \hat{v}(k)$. To be more precise, the proportion

$$q_g := \mathbb{P}(\text{model is correct} \mid G(k) = v_2(n_k) - \hat{v}(k) = g)$$

for $g = 0$ moves from 23.4% at epoch 40 to 76.9% at epoch 210, see Figure 5.13. Additionally for $g = 1$, q_g moves from 61.7% at epoch 40 to 74% at epoch 210. Here the high proportion at the first plateau can be explained from the fact the model learns to approximate $\hat{P}(k)$, which coincides with $\hat{v}(k) + 1$ (i.e. $g = 1$ here) quite often. For higher values of g , these improvements are much smaller, see Table 5.1.

The optimal learning with $\text{factor}(\Delta_k)$ as a feature should be interpreted in light of equations (3.9) and (3.11), as the factors of the discriminant encode not only the primes of bad reduction, but they also make a distinction between those dividing k and those dividing $k^2 - 16$. Among the other experiments that were considered, providing the factors of the conductor $\text{factor}(N_k)$ as a feature leads to substantial learning, but less than giving $\text{factor}(\Delta_k)$. We interpret this phenomenon as providing the primes of bad reduction, but not making the distinction among those that divide k from those that divide $k^2 - 16$. Giving both $\text{factor}(N_k)$ and $\text{factor}(\Delta_k)$ together as features leads to slightly worse learning than giving $\text{factor}(\Delta_k)$ alone, but significantly better than giving $\text{factor}(N_k)$ alone. We speculate that the model has some trouble when provided with too much information.

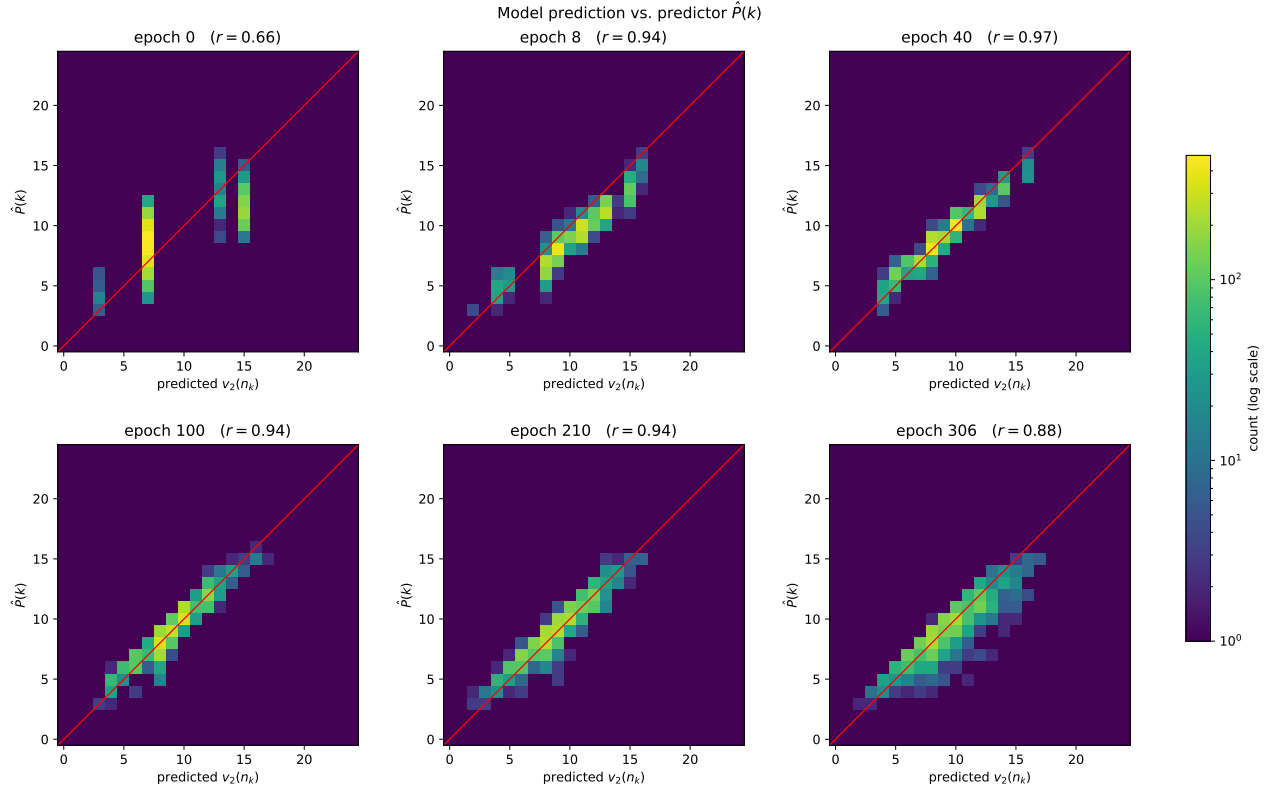
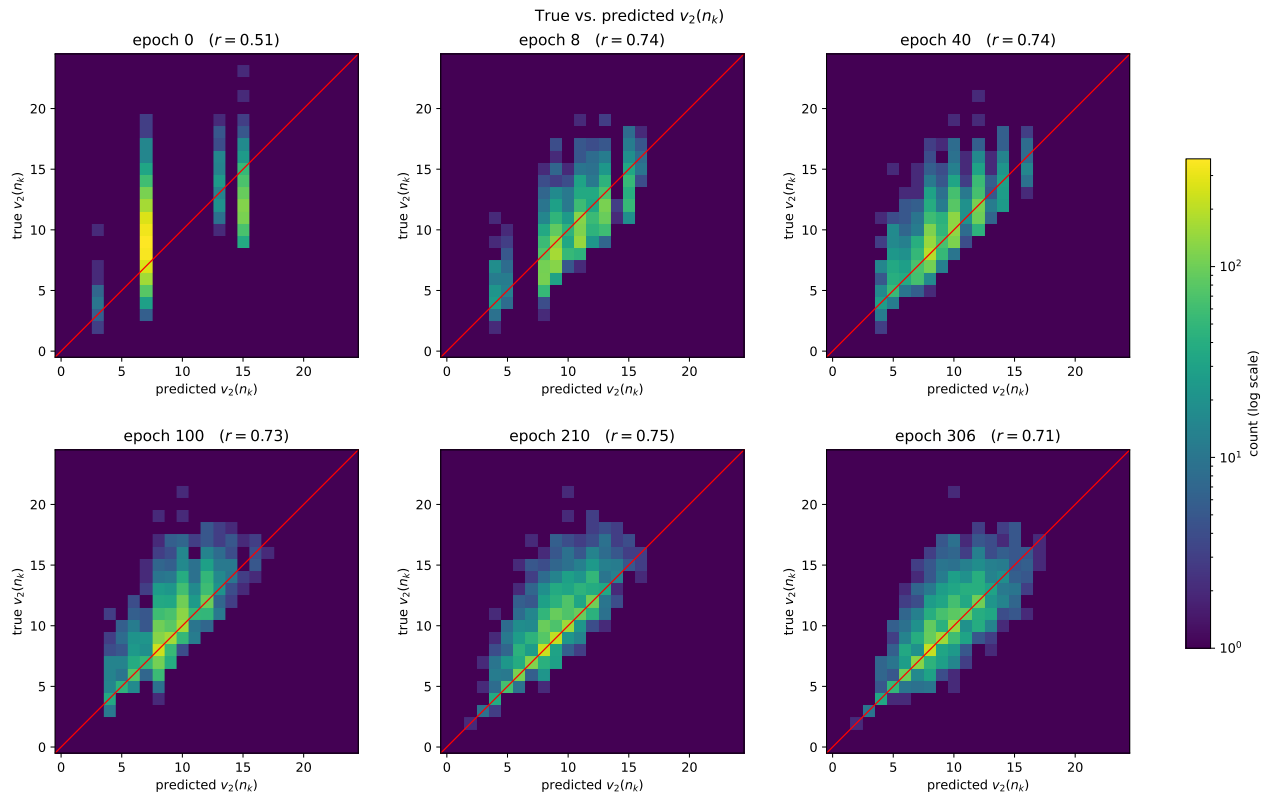
(A) Model prediction vs. predictor $\hat{P}(k) = \text{round}(\hat{v}(k) + s(k) + 1)$.(B) Model prediction vs. true $v_2(n_k)$.

FIGURE 5.11. Confusion-matrix snapshots across training epochs, including the Pearson correlation.

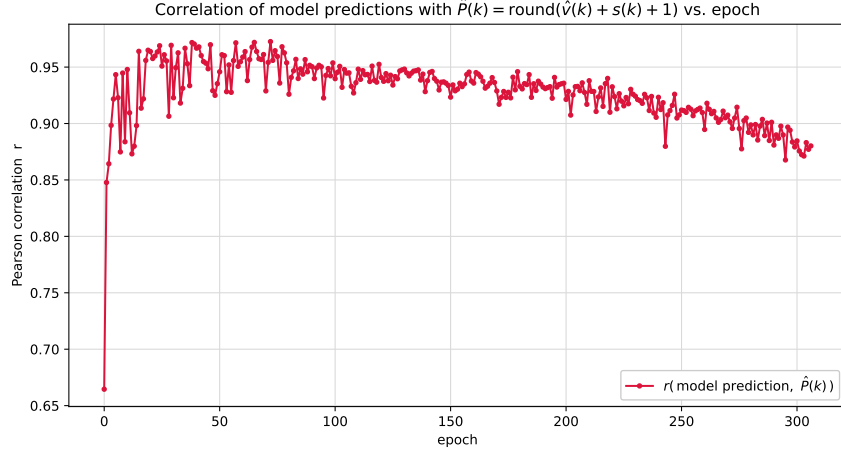


FIGURE 5.12. The Pearson correlation between the model predictions and our predictor $\hat{P}(k)$ for k in the validation set as a function of the epochs.

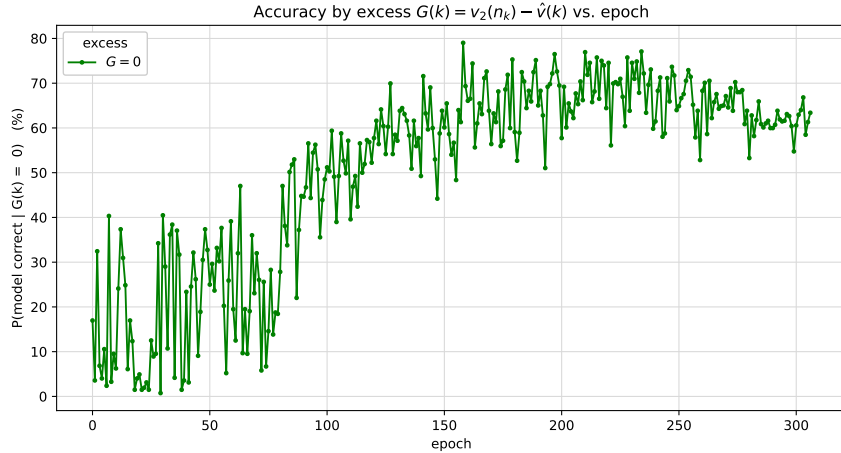


FIGURE 5.13. The proportion of times the model is correct, given that $v_2(n_k) = \hat{v}(k)$ as a function of the epochs.

We found no substantial learning when introducing c_2 , N_k , or Δ_k as features. For the last two cases, this is consistent with the difficulty in factorization.

6. CONCLUSIONS AND FUTURE DIRECTIONS

In this work, we have investigated the rational factors r_k appearing conjecturally in Boyd's identities (0.1) by combining a large-scale numerical study with machine-learning experiments. Rather than revealing a single formula for r_k , our explorations reveal several layers of arithmetic structure. The sign of r_k is governed by the root number, while the conductor determines the order of magnitude of $n_k = 1/r_k$. At the same time, the divisibility properties of n_k are strongly influenced by the local arithmetic of E_k .

One of the most striking conclusions of the statistical analysis is the sharp difference between the behavior at different primes. For primes $p \geq 5$, the positive valuations appear to follow the remarkably simple law $\mathbb{P}(v_p(n_k) = r) = p^{-r}$ for $r \geq 1$, although the deviation observed at $p = 5$

g	$\mathbb{P}(G(k) = g)$	$\mathbb{P}(\text{correct} \mid G(k) = g)$	g	$\mathbb{P}(G(k) = g)$	$\mathbb{P}(\text{correct} \mid G(k) = g)$
0	22.4%	23.4%	0	22.4%	76.9%
1	24.5%	61.7%	1	24.5%	74.0%
2	21.3%	21.5%	2	21.3%	17.9%
3	13.5%	0.2%	3	13.5%	3.5%
4	8.4%	0.0%	4	8.4%	0.4%

(A) Epoch 40 (first plateau). (B) Epoch 210 (second plateau).

TABLE 5.1. The proportion of times the model guesses correctly, under the assumption that the value $G(k) = v_2(n_k) - \hat{v}(k)$ is fixed, together with the distribution of $G(k)$. Comparing the first plateau (A) with the second plateau (B). This is for all k in the validation set.

suggests that this model may require some refinement. The primes 2 and 3 exhibit considerably richer arithmetic. At 3, congruence conditions on k interact with the parity of the number of bad primes congruent to 1 modulo 3, producing laws that are essentially deterministic on large subfamilies. At 2, the valuation is well approximated by a weighted count of the odd primes dividing k and $k^2 - 16$, together with corrections arising from the reduction at 2 and from the 2-adic proximity of k to ± 8 . After these deterministic contributions are removed, the remaining term is closely modeled by a fixed negative-binomial distribution. Although this last observation remains heuristic, the lower bound of Conjecture 3.2 provides a concrete arithmetic statement suggested by the data.

The machine-learning experiments clarify both the possibilities and the limitations of the present approach. Predicting r_k exactly proved to be too difficult, but the model learned its size very effectively when the conductor, or factorizations from which it can be recovered, were included among the inputs. The experiments therefore confirm that the choice and representation of arithmetic features are crucial: the factorizations of the conductor and discriminant are much more informative for the valuation problems than the corresponding unfactored integers. By contrast, the experiments for $p = 5$ and $p = 7$ collapsed to the most frequent output and did not reveal additional structure, illustrating the difficulty caused by highly unbalanced valuation distributions.

The results at 2 and 3 are more encouraging. For the 3-adic valuation, the model recognizes several of the congruence and parity phenomena found in the statistical analysis and continues to improve even on values of k not covered by the largest families considered in Section 3. More surprisingly, in the 2-adic experiment the evolution of the model appears to have a mathematical interpretation. During the first plateau, its predictions are very close to the deterministic predictor $\hat{P}(k)$ derived from our statistical model. During the second plateau, however, its accuracy increases while its correlation with $\hat{P}(k)$ decreases. In particular, the model becomes substantially better at recognizing when the lower bound $\hat{v}(k) \leq v_2(n_k)$ is sharp. This suggests that the network detects additional arithmetic information that is not yet incorporated into our description of $v_2(n_k)$.

These observations lead to several natural questions. The most important is to find a conceptual explanation for the arithmetic of n_k . This includes explaining its near-integrality, proving or refining the proposed distribution for $v_p(n_k)$ when $p \geq 5$, and understanding the congruence and parity laws at 3. At the prime 2, one would like to prove Conjecture 3.2, characterize the values of k for which equality holds, and explain the apparent negative-binomial distribution of the excess. The conditional formula discussed in Remark 3.5 suggests a possible framework for such an explanation.

It would be interesting to develop arithmetic statistics for the families of Tate-twisted Bloch–Kato Selmer groups and regulator indices occurring in that formula. Existing probabilistic models for ordinary Selmer groups, such as those surveyed by Poonen [Poo18], may provide useful guidance, although the groups appearing here are different and the regulator-index term must also be modeled.

For $p \geq 5$, the proposed law

$$\mathbb{P}(v_p(n_k) = r) = p^{-r}$$

says that, after the first factor of p occurs, the higher valuation follows the usual geometric law. The main question is therefore the increased probability

$$\mathbb{P}(v_p(n_k) \geq 1) = \frac{1}{p-1}.$$

It would be interesting to determine whether this enhancement arises from the local Tamagawa factors, the Bloch–Kato Selmer group, the regulator index, or an interaction among these terms. Explicit approaches to p -adic regulators, such as those developed by Coleman–de Shalit [CdS88], Besser [Bes00], and Asakura–Chida [ACB23], may be useful in investigating the regulator contribution. Ultimately, these phenomena should be related to the regulator interpretations underlying Boyd’s conjectures and to the local arithmetic of the elliptic curves E_k .

The machine-learning results also suggest directions for future experiments. Instead of predicting $v_2(n_k)$ directly, one could train a model to predict the excess $G(k) = v_2(n_k) - \hat{v}(k)$, or simply to classify the cases in which $G(k) = 0$. We attempted to use the model to identify the cases in which $G(k) = 0$, but did not obtain a useful classification. Interpreting such a classifier, testing additional arithmetic features, and searching for subfamilies on which it performs unusually well may lead to a refinement of the present conjectures. For the primes $p \geq 5$, the model tends to collapse to the trivial prediction $v_p(n_k) = 0$, since this is the most common class and the loss function only distinguishes correct from incorrect predictions. A class-balanced loss, which assigns greater weight to the rarer positive valuations, or a probabilistic model predicting the full distribution of $v_p(n_k)$, may help prevent this collapse. A first natural extension, taking $k = \sqrt{m}$ for $m \in \mathbb{Z}$ of either sign, is carried out in Section 3.5: $r_{\sqrt{m}}$ is again rational with near-integral reciprocal and the valuation statistics for $p \neq 3$ extend to this setting, but the finer 3-adic parity law of Section 3.3.1 breaks down. Finding its correct generalization to nonsquare m would be very interesting, as would testing whether these patterns persist in other one-parameter families of Mahler measures associated with elliptic curves.

Thus, although the experiments do not provide a formula for r_k , they show that these rational factors are far from arithmetically unstructured. The combination of large-scale computation, statistical analysis, and machine learning has produced concrete conjectures and has identified arithmetic phenomena that are not evident from the currently known identities alone. The next step is to turn these empirical regularities into arithmetic explanations and, eventually, proofs.

7. APPENDIX: TABLES AND FIGURES

(p, r)	$[0, 25K]$	$[0, 50K]$	$[0, 75K]$	$[0, 100K]$	$[0, 125K]$	$[0, 150K]$	$[0, 175K]$	$[0, 200K]$	$[0, 225K]$	$[0, 250K]$	p^{-r}
(5, 1)	0.209	0.208	0.205	0.205	0.204	0.203	0.202	0.202	0.202	0.202	0.200
(5, 2)	0.044	0.044	0.044	0.044	0.043	0.043	0.043	0.043	0.043	0.043	0.040
(5, 3)	0.009	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.009	0.009	0.008
(5, 4)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
(7, 1)	0.142	0.143	0.144	0.144	0.143	0.142	0.142	0.142	0.142	0.142	0.143
(7, 2)	0.021	0.021	0.020	0.020	0.020	0.020	0.021	0.020	0.021	0.021	0.020
(7, 3)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
(11, 1)	0.088	0.090	0.091	0.090	0.091	0.091	0.091	0.090	0.091	0.091	0.091
(11, 2)	0.008	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.009	0.009	0.008
(11, 3)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
(13, 1)	0.074	0.075	0.076	0.076	0.076	0.076	0.076	0.075	0.076	0.076	0.077
(13, 2)	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
(13, 3)	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(17, 1)	0.059	0.058	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059
(17, 2)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
(19, 1)	0.051	0.052	0.051	0.051	0.052	0.051	0.052	0.052	0.052	0.052	0.053
(19, 2)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
(23, 1)	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043	0.043
(23, 2)	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
(29, 1)	0.036	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.035	0.034
(31, 1)	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.033	0.032
(37, 1)	0.025	0.026	0.026	0.026	0.027	0.027	0.027	0.027	0.027	0.027	0.027

TABLE 7.1. Observed cumulative proportions of k satisfying $v_p(n_k) = r$ for the pairs (p, r) listed in the first column, across closed intervals of the form $[0, x]$, where K denotes 1000, together with the reference value p^{-r} .

(p, r)	$[0, 25K]$	$[0, 50K]$	$[0, 75K]$	$[0, 100K]$	$[0, 125K]$	$[0, 150K]$	$[0, 175K]$	$[0, 200K]$	$[0, 225K]$	$[0, 250K]$	p^{-r}
(3, 1)	0.325	0.322	0.323	0.322	0.322	0.321	0.322	0.322	0.321	0.322	0.333
(3, 2)	0.231	0.233	0.234	0.234	0.234	0.234	0.234	0.234	0.235	0.234	0.111
(3, 3)	0.103	0.106	0.107	0.108	0.108	0.110	0.110	0.110	0.110	0.110	0.037
(3, 4)	0.037	0.041	0.042	0.042	0.042	0.042	0.042	0.042	0.043	0.043	0.012
(3, 5)	0.014	0.014	0.014	0.014	0.015	0.015	0.015	0.015	0.015	0.015	0.004
(3, 6)	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.001

TABLE 7.2. Observed cumulative proportions of k satisfying $v_3(n_k) = r$ for the pairs $(3, r)$ listed in the first column, across closed intervals of the form $[0, x]$, where K denotes 1000.

(p, r)	$k \bmod 3$	$[0, 25K]$	$[0, 50K]$	$[0, 75K]$	$[0, 100K]$	$[0, 125K]$	$[0, 150K]$	$[0, 175K]$	$[0, 200K]$	$[0, 225K]$	$[0, 250K]$
$(3, 1)$	all	0.718	0.724	0.728	0.728	0.728	0.730	0.731	0.731	0.732	0.732
	0	0.754	0.755	0.757	0.757	0.757	0.758	0.758	0.758	0.758	0.758
	1	0.698	0.708	0.714	0.713	0.713	0.714	0.715	0.717	0.718	0.719
	2	0.700	0.708	0.713	0.715	0.716	0.717	0.718	0.719	0.719	0.720
$(3, 2)$	all	0.392	0.401	0.405	0.406	0.407	0.408	0.409	0.410	0.411	0.411
	0	0.432	0.436	0.440	0.442	0.442	0.444	0.444	0.445	0.444	0.445
	1	0.375	0.385	0.389	0.389	0.389	0.390	0.391	0.393	0.394	0.394
	2	0.370	0.382	0.384	0.388	0.389	0.391	0.391	0.392	0.394	0.393
$(3, 3)$	all	0.161	0.169	0.171	0.172	0.173	0.175	0.175	0.176	0.176	0.176
	0	0.187	0.193	0.194	0.196	0.195	0.196	0.197	0.197	0.198	0.198
	1	0.149	0.156	0.161	0.163	0.163	0.165	0.166	0.166	0.166	0.166
	2	0.148	0.156	0.157	0.159	0.161	0.163	0.163	0.164	0.165	0.165

TABLE 7.3. Observed cumulative proportions of k satisfying $v_3(n_k) \geq r$ for the pairs $(3, r)$ listed in the first column, split according to congruence classes of $k \bmod 3$, across closed intervals of the form $[0, x]$, where K denotes 1000.

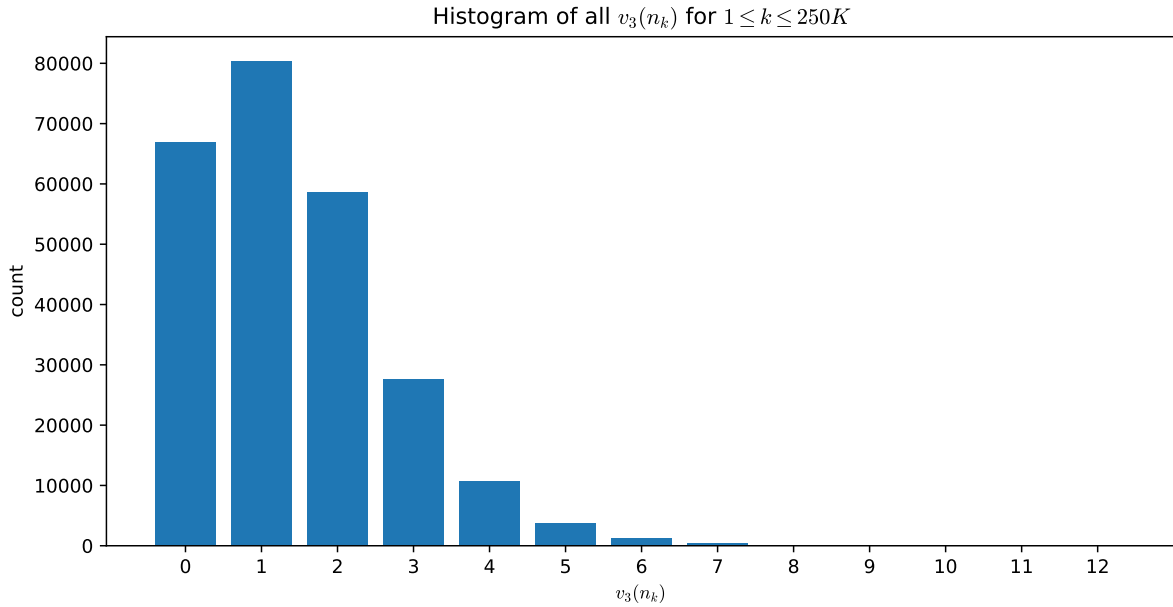


FIGURE 7.1. Histogram of the 3-adic valuations $v_3(n_k)$ for $1 \leq k \leq 250K$ such that $v_3(n_k) \geq 0$, where $K = 1000$.

(p, r)	$[0, 25K]$	$[0, 50K]$	$[0, 75K]$	$[0, 100K]$	$[0, 125K]$	$[0, 150K]$	$[0, 175K]$	$[0, 200K]$	$[0, 225K]$	$[0, 250K]$
(2, 1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(2, 2)	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
(2, 3)	0.011	0.008	0.007	0.006	0.005	0.005	0.005	0.005	0.004	0.004
(2, 4)	0.028	0.021	0.018	0.016	0.015	0.014	0.014	0.013	0.012	0.012
(2, 5)	0.053	0.043	0.038	0.035	0.033	0.031	0.030	0.029	0.028	0.027
(2, 6)	0.089	0.075	0.068	0.064	0.061	0.058	0.056	0.055	0.053	0.052
(2, 7)	0.131	0.115	0.107	0.102	0.097	0.093	0.091	0.089	0.087	0.085
(2, 8)	0.147	0.140	0.133	0.129	0.126	0.123	0.121	0.119	0.117	0.115
(2, 9)	0.148	0.147	0.146	0.144	0.143	0.141	0.140	0.138	0.137	0.136
(2, 10)	0.125	0.134	0.137	0.137	0.137	0.138	0.138	0.138	0.138	0.138
(2, 11)	0.095	0.108	0.114	0.118	0.121	0.123	0.124	0.125	0.125	0.126
(2, 12)	0.067	0.080	0.086	0.090	0.093	0.096	0.098	0.100	0.102	0.103
(2, 13)	0.045	0.053	0.058	0.062	0.065	0.068	0.070	0.072	0.074	0.075
(2, 14)	0.026	0.033	0.037	0.040	0.042	0.044	0.045	0.047	0.048	0.050
(2, 15)	0.015	0.019	0.021	0.024	0.026	0.028	0.029	0.031	0.031	0.032
(2, 16)	0.007	0.010	0.013	0.014	0.015	0.016	0.017	0.018	0.018	0.019
(2, 17)	0.004	0.006	0.007	0.008	0.009	0.009	0.010	0.010	0.011	0.011
(2, 18)	0.003	0.003	0.004	0.005	0.005	0.005	0.005	0.006	0.006	0.006
(2, 19)	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003
(2, 20)	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002

TABLE 7.4. Observed cumulative proportions of k satisfying $v_2(n_k) = r$ for the pairs $(2, r)$ listed in the first column, across closed intervals of the form $[0, x]$, where K denotes 1000.

(p, r)	$k \bmod 7$	$[0, 25K]$	$[0, 50K]$	$[0, 75K]$	$[0, 100K]$	$[0, 125K]$	$[0, 150K]$	$[0, 175K]$	$[0, 200K]$	$[0, 225K]$	$[0, 250K]$
(2, 6)	all	0.905	0.926	0.935	0.941	0.945	0.948	0.951	0.953	0.954	0.956
	0	0.921	0.938	0.947	0.952	0.956	0.958	0.961	0.963	0.964	0.965
	1	0.863	0.892	0.905	0.914	0.919	0.923	0.927	0.930	0.933	0.935
	2	0.870	0.896	0.908	0.914	0.919	0.923	0.927	0.930	0.932	0.934
	3	0.971	0.982	0.987	0.989	0.990	0.991	0.992	0.992	0.993	0.993
	4	0.972	0.981	0.985	0.988	0.989	0.990	0.991	0.991	0.992	0.992
	5	0.865	0.896	0.909	0.916	0.921	0.926	0.928	0.931	0.933	0.935
	6	0.872	0.896	0.908	0.916	0.922	0.925	0.928	0.931	0.933	0.935
(2, 7)	all	0.816	0.851	0.867	0.877	0.884	0.890	0.894	0.898	0.901	0.903
	0	0.835	0.870	0.888	0.896	0.902	0.907	0.911	0.914	0.917	0.919
	1	0.754	0.800	0.818	0.833	0.841	0.848	0.854	0.859	0.862	0.865
	2	0.767	0.805	0.823	0.833	0.841	0.848	0.853	0.859	0.863	0.866
	3	0.917	0.942	0.952	0.958	0.962	0.964	0.967	0.969	0.970	0.971
	4	0.925	0.944	0.954	0.959	0.962	0.965	0.967	0.969	0.970	0.972
	5	0.754	0.798	0.818	0.832	0.842	0.849	0.853	0.858	0.862	0.865
	6	0.758	0.799	0.817	0.831	0.841	0.847	0.853	0.857	0.862	0.864

TABLE 7.5. Observed cumulative proportions of k satisfying $v_2(n_k) \geq r$ for the pairs $(2, r)$ listed in the first column, split according to congruence classes of $k \bmod 7$, across closed intervals of the form $[0, x]$, where K denotes 1000.

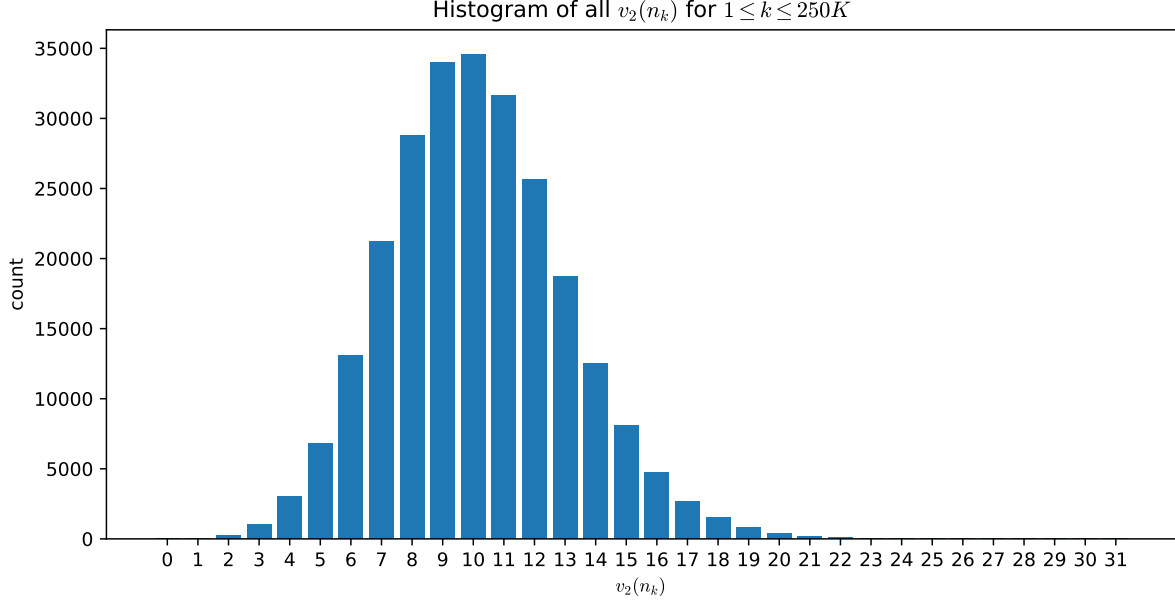


FIGURE 7.2. Histogram of the 2-adic valuations $v_2(n_k)$ for $1 \leq k \leq 250K$ such that $v_2(n_k) \geq 0$, where $K = 1000$.

k	$k - 4$	$k + 4$	$W(k)$	$\hat{v}(k)$	$v_2(n_k)$	$v_2(n_k) - \hat{v}(k)$
321559	$3 \cdot 5 \cdot 13 \cdot 17 \cdot 97$	$11 \cdot 23 \cdot 31 \cdot 41$	21	17	19	+2
400201	$3 \cdot 7 \cdot 17 \cdot 19 \cdot 59$	$5 \cdot 13 \cdot 47 \cdot 131$	21	17	20	+3
440891	$23 \cdot 29 \cdot 661$	$3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 19$	21	17	19	+2
471709	$3 \cdot 5 \cdot 13 \cdot 41 \cdot 59$	$11 \cdot 19 \cdot 37 \cdot 61$	21	17	21	+4
474415	$3 \cdot 7 \cdot 19 \cdot 29 \cdot 41$	$11 \cdot 17 \cdot 43 \cdot 59$	21	17	18	+1
519589	$3 \cdot 5 \cdot 11 \cdot 47 \cdot 67$	$19 \cdot 23 \cdot 29 \cdot 41$	21	17	20	+3
572029	$3 \cdot 5^2 \cdot 29 \cdot 263$	$7 \cdot 11 \cdot 17 \cdot 19 \cdot 23$	21	17	22	+5
729491	$11 \cdot 17 \cdot 47 \cdot 83$	$3^2 \cdot 5 \cdot 13 \cdot 29 \cdot 43$	21	17	21	+4
889291	$3 \cdot 7 \cdot 17 \cdot 47 \cdot 53$	$5 \cdot 11 \cdot 19 \cdot 23 \cdot 37$	23	19	21	+2
1008011	$7 \cdot 11 \cdot 13 \cdot 19 \cdot 53$	$3 \cdot 5 \cdot 17 \cdot 59 \cdot 67$	22	18	20	+2
1394701	$3 \cdot 17 \cdot 23 \cdot 29 \cdot 41$	$5 \cdot 13 \cdot 43 \cdot 499$	22	18	21	+3
1601149	$3^2 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 23$	$83 \cdot 101 \cdot 191$	22	18	21	+3
1847751	$11 \cdot 17 \cdot 41 \cdot 241$	$5 \cdot 7 \cdot 13 \cdot 31 \cdot 131$	22	18	23	+5
1944939	$5 \cdot 19 \cdot 59 \cdot 347$	$7 \cdot 11 \cdot 13 \cdot 29 \cdot 67$	22	18	22	+4

TABLE 7.6. Prime-rich odd values of k in $2.5 \times 10^5 < k < 2 \times 10^6$ chosen to maximise $W(k) := \omega_{\text{odd}}(k) + 2\omega_{\text{odd}}(k^2 - 16)$; the maximum $W = 23$ is attained uniquely at $k = 889291$. For odd k the curve E_k is semistable at 2, so $c(k) = 4$ and $\hat{v}(k) = W(k) - 4$. The table includes every odd k in the window with $W \geq 22$, together with a selection of those with $W = 21$. In every case $v_2(n_k) \geq \hat{v}(k)$.

k	$k - 4$	$k + 4$	$W(k)$	$\hat{v}(k)$	$v_2(n_k)$	$v_2(n_k) - \hat{v}(k)$
940474	$2 \cdot 3 \cdot 5 \cdot 23 \cdot 29 \cdot 47$	$2 \cdot 7 \cdot 11 \cdot 31 \cdot 197$	21	19	21	+2
1069814	$2 \cdot 5 \cdot 7 \cdot 17 \cdot 29 \cdot 31$	$2 \cdot 3 \cdot 37 \cdot 61 \cdot 79$	21	19	27	+8
1195674	$2 \cdot 5 \cdot 7 \cdot 19 \cdot 29 \cdot 31$	$2 \cdot 11 \cdot 17 \cdot 23 \cdot 139$	21	19	20	+1
1204606	$2 \cdot 3 \cdot 7 \cdot 23 \cdot 29 \cdot 43$	$2 \cdot 5 \cdot 11 \cdot 47 \cdot 233$	21	19	22	+3
1234874	$2 \cdot 5 \cdot 7 \cdot 13 \cdot 23 \cdot 59$	$2 \cdot 3 \cdot 29 \cdot 47 \cdot 151$	21	19	22	+3
1258786	$2 \cdot 3 \cdot 7 \cdot 17 \cdot 41 \cdot 43$	$2 \cdot 5 \cdot 13 \cdot 23 \cdot 421$	21	19	21	+2
1286666	$2 \cdot 13 \cdot 17 \cdot 41 \cdot 71$	$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 557$	21	19	21	+2
1540136	$2^2 \cdot 11 \cdot 17 \cdot 29 \cdot 71$	$2^2 \cdot 3 \cdot 5 \cdot 7 \cdot 19 \cdot 193$	21	19	19	+0
1587226	$2 \cdot 3^3 \cdot 7 \cdot 13 \cdot 17 \cdot 19$	$2 \cdot 5 \cdot 23 \cdot 67 \cdot 103$	21	19	20	+1
1614826	$2 \cdot 3 \cdot 11 \cdot 43 \cdot 569$	$2 \cdot 5 \cdot 7 \cdot 17 \cdot 23 \cdot 59$	21	19	20	+1
1653526	$2 \cdot 3 \cdot 13 \cdot 17 \cdot 29 \cdot 43$	$2 \cdot 5 \cdot 37 \cdot 41 \cdot 109$	21	19	22	+3
1669906	$2 \cdot 3 \cdot 13 \cdot 79 \cdot 271$	$2 \cdot 5 \cdot 11 \cdot 17 \cdot 19 \cdot 47$	21	19	20	+1
1700846	$2 \cdot 11 \cdot 13 \cdot 19 \cdot 313$	$2 \cdot 3 \cdot 5^2 \cdot 17 \cdot 23 \cdot 29$	21	19	21	+2
1703464	$2^2 \cdot 3 \cdot 5 \cdot 11 \cdot 29 \cdot 89$	$2^2 \cdot 13 \cdot 17 \cdot 41 \cdot 47$	21	19	26	+7
1717166	$2 \cdot 41 \cdot 43 \cdot 487$	$2 \cdot 3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 37$	21	19	24	+5
1722886	$2 \cdot 3 \cdot 7 \cdot 17 \cdot 19 \cdot 127$	$2 \cdot 5 \cdot 13 \cdot 29 \cdot 457$	21	19	21	+2
1762094	$2 \cdot 5 \cdot 11 \cdot 83 \cdot 193$	$2 \cdot 3 \cdot 13 \cdot 19 \cdot 29 \cdot 41$	21	19	24	+5
1774630	$2 \cdot 3 \cdot 7 \cdot 29 \cdot 31 \cdot 47$	$2 \cdot 23 \cdot 173 \cdot 223$	21	19	22	+3
1791566	$2 \cdot 17 \cdot 23 \cdot 29 \cdot 79$	$2 \cdot 3 \cdot 5 \cdot 11 \cdot 61 \cdot 89$	21	19	24	+5
1812794	$2 \cdot 5 \cdot 7 \cdot 19 \cdot 29 \cdot 47$	$2 \cdot 3^2 \cdot 13 \cdot 61 \cdot 127$	21	19	19	+0
1885594	$2 \cdot 3^2 \cdot 5 \cdot 7 \cdot 41 \cdot 73$	$2 \cdot 11 \cdot 13 \cdot 19 \cdot 347$	21	19	21	+2
1899206	$2 \cdot 19 \cdot 23 \cdot 41 \cdot 53$	$2 \cdot 3 \cdot 5 \cdot 29 \cdot 37 \cdot 59$	21	19	21	+2
1952786	$2 \cdot 13 \cdot 19 \cdot 59 \cdot 67$	$2 \cdot 3 \cdot 5 \cdot 7 \cdot 17 \cdot 547$	21	19	20	+1
1984814	$2 \cdot 5 \cdot 41 \cdot 47 \cdot 103$	$2 \cdot 3 \cdot 11 \cdot 17 \cdot 29 \cdot 61$	21	19	19	+0

TABLE 7.7. Prime-rich even values of k in $2.5 \times 10^5 < k < 2 \times 10^6$ chosen to maximise $W(k) := \omega_{\text{odd}}(k) + 2\omega_{\text{odd}}(k^2 - 16)$; for even k the maximum is $W = 21$. All are even with $16 \nmid k$, so E_k is additive at 2, $c(k) = 2$ and $\hat{v}(k) = W(k) - 2$. The table lists every even k in the window attaining $W = 21$, i.e. the complete set of even maximisers. In every case $v_2(n_k) \geq \hat{v}(k)$, with equality at $k = 1540136$, $k = 1812794$, and $k = 1984814$.

REFERENCES

- [ACB23] Masanori Asakura, Masataka Chida, and François Brunault. A numerical approach toward the p -adic Beilinson conjecture for elliptic curves over $\mathbb{Q}$. *Res. Math. Sci.*, 10(1):Paper No. 11, 58, 2023.
- [Bes00] Amnon Besser. Syntomic regulators and p -adic integration. II. K_2 of curves. In *Proceedings of the Conference on p -adic Aspects of the Theory of Automorphic Representations (Jerusalem, 1998)*, volume 120, pages 335–359, 2000.
- [Boy98] David W. Boyd. Mahler's measure and special values of L -functions. *Experiment. Math.*, 7(1):37–82, 1998.
- [Bru06] François Brunault. Version explicite du théorème de Beilinson pour la courbe modulaire $X_1(N)$. *C. R. Math. Acad. Sci. Paris*, 343(8):505–510, 2006.
- [Bru16] François Brunault. Regulators of Siegel units and applications. *J. Number Theory*, 163:542–569, 2016.
- [CdS88] Robert Coleman and Ehud de Shalit. p -adic regulators on curves and special values of p -adic L -functions. *Invent. Math.*, 93(2):239–266, 1988.
- [Cha25] François Charton. Int2int: a framework for mathematics with transformers. *arXiv:2502.17513 [cs.LG]*, 2025.
- [Den97] Christopher Deninger. Deligne periods of mixed motives, K -theory and the entropy of certain $\mathbf{Z}^n$ -actions. *J. Amer. Math. Soc.*, 10(2):259–281, 1997.
- [DGdJK26] Neil Dummigan, Vasily Golyshev, Rob de Jeu, and Matt Kerr. The 2-part of the bloch-kato conjecture, and indivisibility results, for K_2 of some elliptic curves. *arXiv preprint arXiv:2605.11100*, 2026.
- [KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [Lal10] Matilde N. Lalín. On a conjecture by Boyd. *Int. J. Number Theory*, 6(3):705–711, 2010.
- [LR07] Matilde N. Lalín and Mathew D. Rogers. Functional equations for Mahler measures of genus-one curves. *Algebra Number Theory*, 1(1):87–117, 2007.
- [LSZ16] Matilde N. Lalín, Detchat Samart, and Wadim Zudilin. Further explorations of Boyd's conjectures and a conductor 21 elliptic curve. *J. Lond. Math. Soc. (2)*, 93(2):341–360, 2016.
- [Mel11] Anton Mellit. Mahler measures and q -series. *Oberwolfach Reports*, 8(3):1990–1991, 2011. Abstract from the workshop “Explicit Methods in Number Theory,” Oberwolfach, Germany, 17–23 July 2011.
- [Mel19] Anton Mellit. Elliptic dilogarithms and parallel lines. *J. Number Theory*, 204:1–24, 2019.
- [PAR25] PARI Group, Univ. Bordeaux. *PARI/GP version 2.17.2*, 2025. Available from <http://pari.math.u-bordeaux.fr/>.
- [Poo18] Bjorn Poonen. Heuristics for the arithmetic of elliptic curves. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. II. Invited lectures*, pages 399–414. World Sci. Publ., Hackensack, NJ, 2018.
- [RV99] Fernando Rodriguez-Villegas. Modular Mahler measures. I. In *Topics in number theory (University Park, PA, 1997)*, volume 467 of *Math. Appl.*, pages 17–48. Kluwer Acad. Publ., Dordrecht, 1999.
- [RZ12] Mathew D. Rogers and Wadim Zudilin. From L -series of elliptic curves to Mahler measures. *Compos. Math.*, 148(2):385–414, 2012.
- [RZ14] Mathew D. Rogers and Wadim Zudilin. On the Mahler measure of $1 + X + 1/X + Y + 1/Y$. *Int. Math. Res. Not. IMRN*, (9):2305–2326, 2014.
- [VSP⁺17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [Zud14] Wadim Zudilin. Regulator of modular units and Mahler measures. *Math. Proc. Cambridge Philos. Soc.*, 156(2):313–326, 2014.

ALBERTO ALFARANO: AXIOM MATH
 Email address: alberto@axiommath.ai

PABLO BIANUCCI: DEPARTMENT OF PHYSICS, CONCORDIA UNIVERSITY, 7141 SHERBROOKE ST. W. MONTREAL, QC H4B 1R6, CANADA
 Email address: pablo.bianucci@concordia.ca

MATILDE LALÍN: DÉPARTEMENT DE MATHÉMATIQUES ET DE STATISTIQUE, UNIVERSITÉ DE MONTRÉAL. CP 6128, SUCC. CENTRE-VILLE. MONTREAL, QC H3C 3J7, CANADA

Email address: `matilde.lalin@umontreal.ca`

BEREND RINGELING: DÉPARTEMENT DE MATHÉMATIQUES ET DE STATISTIQUE, UNIVERSITÉ DE MONTRÉAL. CP 6128, SUCC. CENTRE-VILLE. MONTREAL, QC H3C 3J7, CANADA

Email address: `bjringeling@gmail.com`