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 observed frequencies of $v_p(r_k) = -m$, for $m \geq 1$, are close to p^{-m}. At $p = 5$, the agreement improves substantially when we restrict to curves with good reduction at 5. 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 predictive relationships within the computed range that are not fully captured by the explicit predictor obtained from our statistical analysis.

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

$$\mathrm{m}(P) = \frac{1}{(2\pi i)^n} \int_{\mathbb{T}^n} \log |P(x_1, \ldots, x_n)| \, \frac{\mathrm{d}x_1}{x_1} \cdots \frac{\mathrm{d}x_n}{x_n},$$

where $\mathbb{T}^n = \{(u_1, \ldots, u_n) \in \mathbb{C}^n : |u_1| = \cdots = |u_n| = 1\}$. 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 $\mathrm{m}(P) = \log |a| + \sum_j \log \max\{1, |\alpha_j|\}$.

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 Beĭlinson'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

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

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 $k \in \mathbb{Z} \setminus \{-4, 0, 4\}$,

$$\mu_k = r_k L'(E_k, 0), \tag{0.1}$$

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 because it enters the functional equation of the L-function and governs the scale of $L'(E_k, 0)$.

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 $x \to -x$ and $y \to -y$ 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].

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

2020 *Mathematics Subject Classification.* Primary 11R06; Secondary 11F67, 11-11, 68T05.

Key words and phrases. Mahler measure, elliptic curves, machine learning, L-functions.

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 $1/r_k$ and the conductor of E_k.

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 Beĭlinson'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 $p \geq 5$, we compare the observed positive valuations with the reference probabilities p^{-r} for $r \geq 1$. This provides a useful description for several primes over the computed range. At $p = 5$, the agreement improves substantially when we restrict to curves with good reduction at 5.

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 $1 \bmod 3$. For the 2-adic valuation, the data suggest the lower bound

$$v_2(n_k) \geq \omega_{\mathrm{odd}}(k) + 2\,\omega_{\mathrm{odd}}(k^2 - 16) - c(k), \qquad k \neq 4, 8,$$

where $\omega_{\mathrm{odd}}(n)$ counts the distinct odd prime divisors of n, and $c(k)$ is the correction term defined in (3.8). The excess over this proposed lower bound also depends on the 2-adic class of k. In particular, the empirical class means are approximately described by $2 + s(k)$, with $s(k)$ defined in (3.10). 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 frequent zero class provides a strong baseline. Some runs at $p = 5$ correctly predict positive valuations, but their overall validation accuracy at the final epoch is below that of the constant-zero predictor. 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 improved accuracy at the second plateau suggests predictive relationships within the computed range that are not fully captured by our explicit predictor $\hat{P}(k)$ (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, Rob de Jeu, Dimitris Koukoulopoulos, 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 used to assist with processing and cleaning the dataset and to suggest arithmetic patterns. These pattern searches were exploratory; the resulting observations and conjectures are distinguished from the proved statements.
- 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.
- Computational Checks and Editing: OpenAI's ChatGPT was used to assist with computational checks, analysis scripts, model evaluation, figure preparation, and improvements to 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 and datasets used in this study are publicly available through Borealis, the Canadian Dataverse Repository [ABLR26].

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

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

with inverse

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

which leads to the Weierstrass equation

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

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 main 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 $|m| \le B$, 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]):

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

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 PARI/GP at approximately 19 digits of internal precision. We then applied `lindep` to the vector $(\mu_k, L'(E_k, 0))$, without specifying an additional precision parameter. If an integer relation

$$A\mu_k + BL'(E_k, 0) \approx 0, \qquad A \ne 0,$$

was returned, we recorded the rational candidate $r_k = -B/A$. No additional residual threshold or bound on the coefficients was imposed. In our dataset, the largest absolute numerator of r_k is 11, and the largest denominator is 40,167,578,001,408. The numerator has absolute value 1 in 249,993 of the 249,999 cases, and the median denominator is 299,186,683,904.

To assess the reliability of this rational reconstruction, we independently recomputed μ_k and $L'(E_k, 0)$ with precision targets of 50 and 100 decimal digits (and 30 additional guard digits in each case) for a reproducibly selected random sample of 30 values of k from $1 \le k \le 250{,}000$, $k \ne 4$. More precisely, the computed relative discrepancy

$$\frac{|\mu_k - r_k L'(E_k, 0)|}{|\mu_k|}$$

was below 2×10^{-96} in the computations with requested precision 50 decimal digits and below 10^{-134} in those with requested precision 100 decimal digits. The verification code and the selected values of k are available in our code repository.

We computed the values r_k for $1 \le k \le 250{,}000$ (with $k \ne 4$). 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 rational values r_k numerically recognized by the procedure described in Section 2, for $1 \le k \le 250{,}000$ with $k \ne 4$. Since $\mu_k > 0$, the functional equation of the L-function determines the sign of r_k from the root number ω_k of E_k.

We write $n_k := 1/r_k$, including the cases in which n_k is not an integer. In the computed dataset, n_k is an integer whenever

$$k \notin \{3, 4, 5, 8, 12, 16, 32\}. \tag{3.1}$$

We examine the sizes and p-adic valuations of these values n_k below.

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

$$\mu_k = \log(k) + o(1) \tag{3.2}$$

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

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

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*

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

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

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

where

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

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 $p \| N$, 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 $p^2 | N$, 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

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

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.

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.

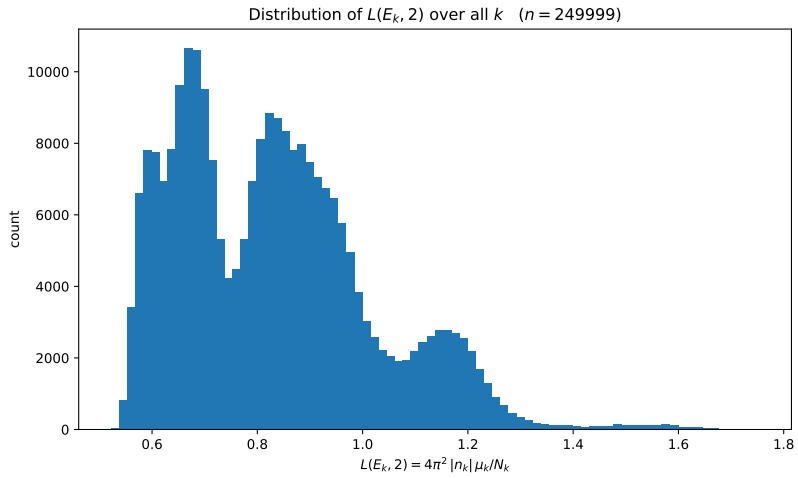


FIGURE 3.1. Values of $L(E_k, 2)$ for $k \leq 250{,}000$ in our dataset.

3.2. **p-adic observations for primes $p \geq 5$.** Over the computed range, the observed frequencies suggest the approximation

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

compared with the probability $1/p$ predicted by a uniform-distribution model modulo p. Conditional on a positive valuation, the observed frequencies are close to the usual geometric distribution:

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

Together, these observations suggest the approximation

$$\mathbb{P}\big(v_p(n_k) = r\big) \approx p^{-r}, \qquad r \geq 1.$$

The distinctive feature of the proposed model is therefore the increased probability of the first factor of p; the conditional distribution of the higher valuations is the usual geometric distribution. 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 $v_5(n_k) \geq 6$. In these comparisons, the first bin is $v_p(n_k) \leq 0$. For $p \geq 5$, this differs from $v_p(n_k) = 0$ only at $p = 11$, $k = 16$, where $n_{16} = 1/11$. The reference probability of this bin is $(p-2)/(p-1)$. If O_i is the observed number of values in bin i, and E_i is the number predicted by the random model, then

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

We also report the normalized statistic χ^2/ν, where

$$\nu = \#\{\text{bins}\} - 1.$$

We use these statistics as descriptive measures of agreement between the observed counts and the model. Their values depend on the sample size, so small differences in the observed frequencies can produce elevated statistics. We interpret these comparisons descriptively, without assuming that the observations indexed by consecutive integers k are independent random samples.

In Table 3.1a, we compare the observed p-adic valuations with the model assigning probability p^{-r} to valuation $r \geq 1$ and probability $(p-2)/(p-1)$ to valuation 0. Throughout this section, $\mathbb{P}$ denotes the empirical proportion obtained by sampling k uniformly from the indicated set of parameters.

For each prime $5 \leq p \leq 47$, we make this comparison both for the full dataset and for the subfamily for which E_k has good reduction at p, equivalently $p \nmid k(k^2 - 16)$. For each sample, we compute the chi-square statistic, the

number of degrees of freedom, and the normalized statistic χ^2/ν. The expected counts are computed using the size of the sample being compared.

For $p = 5$, restricting to good reduction substantially improves the agreement with the model: χ^2/ν decreases from 12.81 to 1.49. The counts for both samples are recorded in Table 3.1b. For $p = 37$, the corresponding normalized statistic decreases from 4.76 to 4.09.

	all k			E_k good at p		
p	χ^2	ν	χ^2/ν	χ^2	ν	χ^2/ν
5	76.84	6	12.81	8.97	6	1.49
7	1.56	5	0.31	3.72	5	0.74
11	8.32	4	2.08	3.34	4	0.83
13	6.42	4	1.60	6.90	4	1.72
17	1.49	3	0.50	2.67	3	0.89
19	4.10	3	1.37	2.37	3	0.79
23	3.37	3	1.12	0.69	3	0.23
29	3.41	3	1.14	3.18	3	1.06
31	7.33	3	2.44	5.26	3	1.75
37	14.27	3	4.76	12.27	3	4.09
41	2.60	2	1.30	3.07	2	1.54
43	2.41	2	1.20	2.37	2	1.18
47	1.38	2	0.69	1.39	2	0.69

(A) Comparison with the model for primes $5 \leq p \leq 47$, using all k and, separately, those k for which E_k has good reduction at p.

	all k			E_k good at 5		
$v_5(n_k) = r$	count	expected	ratio	count	expected	ratio
0	186,238	187,499.3	0.993	75,046	75,000.0	1.001
1	50,398	49,999.8	1.008	19,802	20,000.0	0.990
2	10,651	10,000.0	1.065	4,101	4,000.0	1.025
3	2,141	2,000.0	1.071	825	800.0	1.031
4	471	400.0	1.178	182	160.0	1.138
5	78	80.0	0.975	34	32.0	1.063
6	19	16.0	1.188	9	6.4	1.406
7	3	3.2	0.938	1	1.28	0.781

(B) Counts with $v_5(n_k) = r$ versus the expected values. The left block uses all $M = 249{,}999$ parameters, while the right block uses the $M_{\text{good}} = 100{,}000$ parameters for which E_k has good reduction at 5. For $r \geq 1$, the expected count is $M \cdot 5^{-r}$ (respectively $M_{\text{good}} \cdot 5^{-r}$); for $r = 0$, the expected proportion is $\frac{3}{4}$.

TABLE 3.1. Comparison of the observed p-adic valuations with the model assigning probability p^{-r} to positive valuation r. Left: the χ^2 statistics for the full dataset and the subfamily with good reduction at p. Right: counts for $v_5(n_k) = r$ in the full dataset and in the subfamily with good reduction at 5.

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 $1/3 = 0.333\ldots$ 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 $k \equiv 0 \bmod 3$; see Table 7.3. Refining modulo 27, we find a stronger statement: for $k \equiv 0, \pm 5 \bmod 27$ 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

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

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 $\mathbb{P}(v_3(n_k) \geq 1 \,|\, \#S_k \text{ is even}) = 0.79$, compared to $\mathbb{P}(v_3(n_k) \geq 1 \,|\, \#S_k \text{ is odd}) = 0.67$. If we compare this to the analogous set

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

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

	$v_3(n_k) \le 0$	$v_3(n_k) \ge 1$	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

	$v_3(n_k) \le 0$	$v_3(n_k) \ge 1$	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

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

the above-mentioned correlations sharpen into two implications that hold throughout the computed range $1 \le k \le 250{,}000$, with the exception of $k = 1$:

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

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

$v_3(n_k)$	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 $v_3(n_k) = 0 \Rightarrow \#S_k$ odd and $v_3(n_k) = 1 \Rightarrow \#S_k$ even holds with a single exception; for $v_3(n_k) \ge 2$ the parity is unconstrained.

	$v_3(n_k) = 0$	$v_3(n_k) \ge 1$	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 $v_3(n_k) \ge 1$ for $k \in A$.

The observed laws (3.6) suggest a generalization to larger families. Define

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

The two laws are

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

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

On A we have k^2-16 is squarefree, $v_3(k)=0$ and $8\nmid k$, so (3.7) reduces to (3.6). Table 3.5 and Figure 3.2 compare several enlargements of A. Among the families listed, the largest one with no observed exceptions to the odd law (3.7a) in the range $5\le k\le 250{,}000$ is

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

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 $v_3(n_k)\ge 1$ 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 $1 \bmod 3$. This suggests that both laws (3.7) may hold on the smaller set

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

The additional restriction defining C'' was motivated by the exceptions to the even law on C' within the computed range. Thus the absence of exceptions to either law on C'' in Table 3.5 is exploratory evidence obtained from the same data used to formulate this restriction. Finally, it is worth noting that the k such that $k\equiv 0 \bmod 27$ or $k\equiv \pm 5 \bmod 27$ 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 $3\equiv 3 \bmod 4$. So their elevated 3-valuation is not explained by the law on C''.

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'	$B\setminus\{k\equiv 4 \bmod 8\}$	0.565	0	1,538
B''	$B'\cap\{p^3\mid k\Rightarrow p\equiv 2 \bmod 3\}$	0.536	0	0
C	$p^2\nmid k^2-16$ for all $p\equiv 3 \bmod 4$	0.721	3,423	5,330
C'	$C\setminus\{k\equiv 4 \bmod 8\}$	0.631	0	1,746
C''	$C'\cap\{p^3\mid k(k^2-16)\Rightarrow p\equiv 2 \bmod 3\}$	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 $5\le k\le 250{,}000$. Here "density" is the proportion of integers k that lie in the set, and "exc." is the number of exceptions of the corresponding law.

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 $v_2(n_k)\ge 2$ 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 $k\equiv a \bmod q$ with $v_2(n_k)\ge r$. 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 $a\equiv\pm 4 \bmod q$ and second largest when $a\equiv 0 \bmod q$; see for example Table 7.5, where the frequency of $v_2(n_k)\ge r$ is highest for $k\equiv\pm 4 \bmod 7$, closely followed by $k\equiv 0 \bmod 7$. These are exactly the classes of bad reduction: since $\Delta_k = 2^{-24}k^{14}(k^2-16)$, we have $\Delta_k\equiv 0 \bmod q$ if and only if $k\equiv 0,\pm 4 \bmod q$.

Empirical model. From the earlier observations, the primes dividing k^2-16 (the case $k\equiv\pm 4 \bmod q$) have more influence than those dividing k (the case $k\equiv 0 \bmod q$), and empirically the appropriate weights are 2 and 1 respectively. We therefore assign to each odd prime q the weight

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

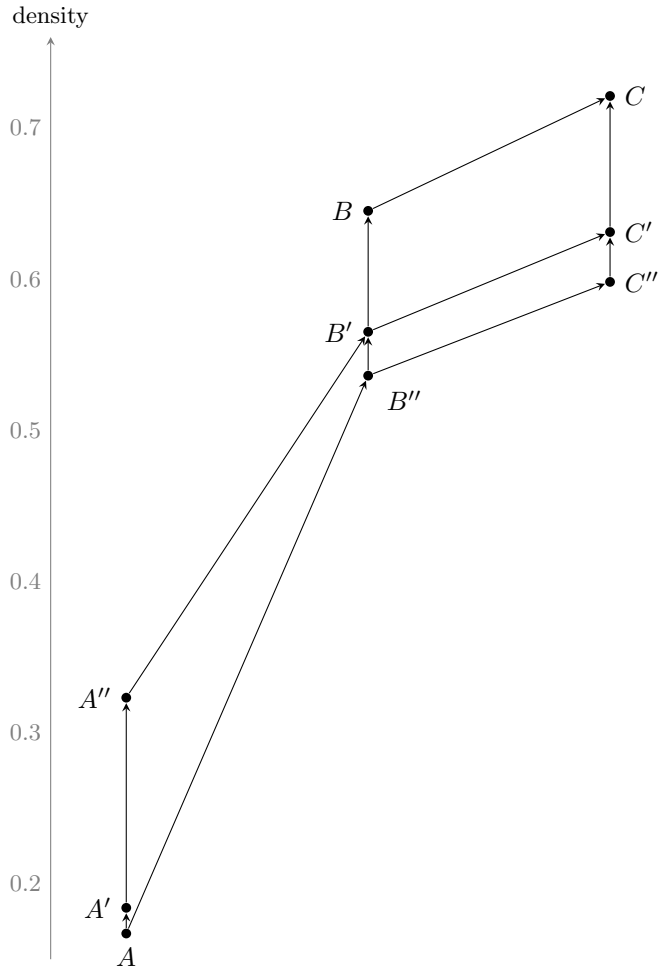


FIGURE 3.2. The sets of Table 3.5, placed at height equal to their density. An arrow points from a set to a larger set containing it; since a subset is never denser than its superset, every arrow points upward.

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

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

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

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

We then set

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

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.

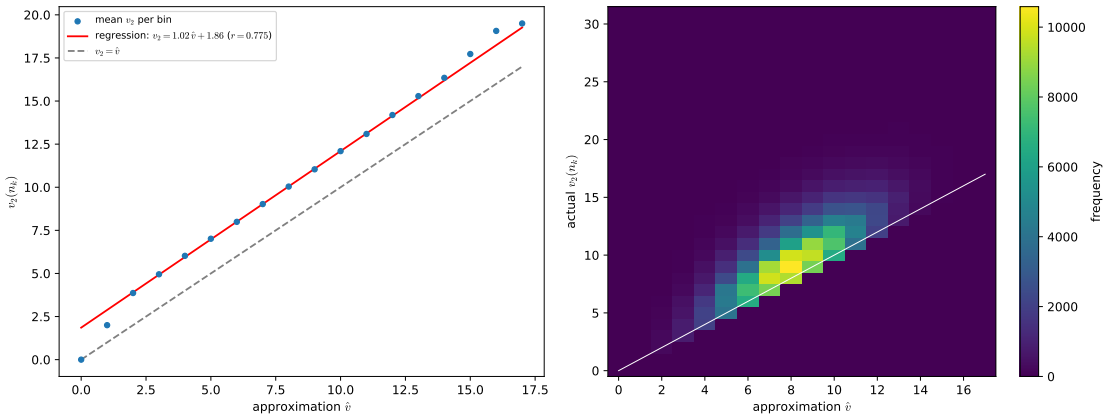


FIGURE 3.3. $v_2(n_k)$ vs $\hat{v}(k)$ for k with $v_2(n_k) \geq 0$

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

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

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 $k \equiv \pm 8 \bmod 64$, where the mean excess is 3.02 (see Table 3.6b). In fact it grows further as $k \equiv \pm 8 \bmod 2^m$ for increasing $m \geq 6$, plateauing at an excess of 3.5 (see Table 3.6c).

A simple approximation to the empirical mean of $G(k)$ within the corresponding 2-adic classes is $2 + s(k)$, where

$$s(k) := \tfrac{3}{2} \max\Big\{1 - 2^{\,5 - \max\{v_2(k-8),\, v_2(k+8)\}},\, 0\Big\}. \tag{3.10}$$

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

For comparison with the overall valuation histogram, we use the integer-valued reference model

$$\hat{v}(k) + \operatorname{round}\big(s(k)\big) + X_k, \qquad X_k \sim \mathrm{NB}\big(3, \tfrac{3}{5}\big)\,, \tag{3.11}$$

where round denotes rounding to the nearest integer and X_k is an auxiliary random variable. Here NB denotes the negative-binomial distribution, with probability mass function

$$\mathbb{P}(X_k = j) = \binom{j+2}{2} \left(\frac{3}{5}\right)^3 \left(\frac{2}{5}\right)^j, \qquad j = 0, 1, 2, \ldots.$$

This distribution has mean 2 and mode 1. The mean excess in the rounded reference model is therefore $2 + \operatorname{round}(s(k))$. The quantity $2 + s(k)$ approximates the observed averages of $G(k)$ listed in Table 3.6c.

Figure 3.4 compares this reference model with the observed distribution of $v_2(n_k)$. We regard it as a first-order description of the overall histogram.

For individual valuations, we use the empirical predictor

$$\hat{P}(k) := \operatorname{round}\big(\hat{v}(k) + s(k) + 1\big). \tag{3.12}$$

On the sample shown in Figure 3.4, this predictor is the mode of the rounded reference model. 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,

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

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	12.34	+3.45	$\to 3.5$

(C) Excess by exact 2-adic depth $M := \max\{v_2(k-8), v_2(k+8)\}$, compared with the approximation $2 + s(k)$ to the observed average excess. The arrow in the last column indicates that $2 + s(k)$ tends to 3.5 as $M \to \infty$.

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 $M = \max\{v_2(k-8), v_2(k+8)\}$.

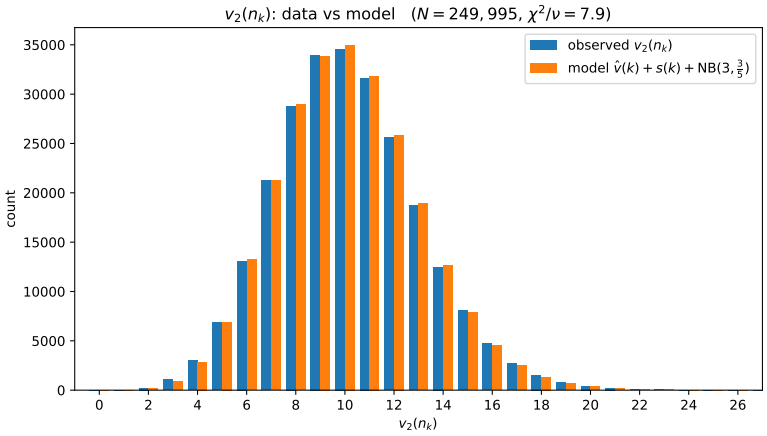


FIGURE 3.4. The observed distribution of $v_2(n_k)$ for values with $v_2(n_k) \geq 0$, compared with the reference model $\hat{v}(k) + \mathrm{round}(s(k)) + \mathrm{NB}(3, \frac{3}{5})$. The shift $s(k)$ appearing in the legend is understood with this nearest-integer discretization.

as $x \to \infty$. 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 report 38 further values in the range $2.5 \times 10^5 < k < 2 \times 10^6$, selected for large values of

$$W(k) := \omega_{\mathrm{odd}}(k) + 2\,\omega_{\mathrm{odd}}(k^2 - 16).$$

These computations provide targeted checks of Conjecture 3.2 beyond the main dataset, at parameters with many odd primes of bad reduction. All 38 values satisfy $v_2(n_k) \geq \hat{v}(k)$, with equality at $k = 1540136$, $k = 1812794$, and $k = 1984814$. The value $k = 1069814$ gives $v_2(n_k) = 27$; the largest 2-adic valuation observed in the main dataset is 31, attained at $k = 36326$.

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

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

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

We have computed these r_k for all integers m with $|m| \leq 125{,}000$, excluding the singular parameters $m = 0$ and $m = 16$. 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 the primes considered in Table 3.1a, the observed frequencies are broadly described by the empirical approximation $\mathbb{P}(v_p(n_k) = r) \approx p^{-r}$ for $r \geq 1$. For $p = 5$, the proportion with $v_5(n_k) \geq 1$ is 26.24%, compared with the reference value 25%. Restricting to the 150,000 parameters for which E_k has good reduction at 5, equivalently $5 \nmid m(m-16)$, gives 26.84%. Thus the excess in the proportion of positive 5-valuations persists under this restriction.

For $p = 3$, the parity relation (3.6) is satisfied for only about half of the parameters for which $v_3(n_k) \in \{0, 1\}$. Among the residue classes modulo 27, the class $m \equiv 25 \bmod 27$ is the only one for which every computed value with integral n_k satisfies $v_3(n_k) \geq 1$. This class is the image of $k \equiv \pm 5 \bmod 27$.

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

$$\omega^0_{\mathrm{odd}}(m) + 2\omega^1_{\mathrm{odd}}(m),$$

where $\omega^0_{\mathrm{odd}}(m)$ is the number of odd primes p dividing m with $v_p(m)$ even, and $\omega^1_{\mathrm{odd}}(m)$ 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

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

where

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

Then $\hat{v}(m) \leq v_2(n_k)$ appears to hold for all m in our data for which n_k is integral, the sole exception being $m = 12$.

Equality is attained for 40,099 of the 249,986 parameters for which n_k is integral, or about 16%.

Apart from the stated exception, the residual $v_2(n_k) - \hat{v}(m)$ is non-negative, and its observed distribution is of the same negative-binomial flavor as that of the excess $G(k)$ 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

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

The curve $\widetilde{E}_u$ 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 $u \equiv 4 \bmod 8$ and $u(u^2-1)$ is squarefree at the odd primes; assume in addition that $u+1$ has a prime divisor $\equiv 3 \bmod 4$. Assume moreover, as in* [DGdJK26, Thm. 3]*, that $K_2^T(\tilde{E}_u)_{\mathrm{int}} \otimes_{\mathbb{Z}} \mathbb{Q}$ is one-dimensional, that r_k is a rational number and that $H^1_f(\mathbb{Q}, \tilde{E}_u[2^\infty](-1))$ 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

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

By the functional equation and the fact that $\widetilde{E}_u$ 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

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

where

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

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

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

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

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

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 $1 \bmod 4$. 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

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

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 $q \equiv 1 \bmod 4$, 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. □

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

$$\begin{aligned} v_p(n_k) = v_p(N_k) &+ \sum_{q\le\infty} v_p\Big(\mathrm{Tam}^0_{q,\omega}(H^1_{\text{ét}}(\widetilde{E}_{u,\overline{\mathbb{Q}}},\mathbb{Z}_p)(2))\Big) \\ &+ v_p\Big(\#H^1_f(\mathbb{Q},\widetilde{E}_u[p^\infty](-1))\Big) \\ &- v_p\Big(\#H^0(\mathbb{Q},\widetilde{E}_u[p^\infty](1))\Big) - v_p\Big(\#H^0(\mathbb{Q},\widetilde{E}_u[p^\infty](-1))\Big) \\ &- v_p(\iota_{u,p}), \end{aligned}$$

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.

For the principal experiments, we randomly split our set into *training*, *validation* and *test sets*, of sizes 243,999, 3,000 and 3,000, respectively. We also consider two blocked splits, described in Section 5.4.2. The training set is used to train the neural network. The validation set is used to monitor learning over the course of training and to select the checkpoints reported in the discussion. We trained each model from scratch on the training split. At the end of each epoch, the evaluation pipeline computed and archived predictions on both the validation and test splits. Checkpoint selection used only validation accuracy; the test split was not used for model or checkpoint selection. We report the archived test scores corresponding to the validation-selected checkpoints.

The learning curves use the validation split. Greedy accuracy is the percentage of predictions obtained by greedy decoding that are correct under the task's scoring criterion. Each epoch consists of 1,000 training steps, and epoch labels start at 0.

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 in base 10 (as a default), but the tasks involving the learning of $v_p(n_k)$ are performed in base p, since this allows the model to detect multiplicity by p and the value of v_p more easily than in other bases.

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 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 factor(Δ_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 bounds for E_k, computed using `ellrank` in PARI/GP [PAR25]. We use the lower and upper bounds, represented as the vector $[\mathrm{rank}_{\min}, \mathrm{rank}_{\max}]$. These bounds determine the rank when they coincide. The Birch and Swinnerton-Dyer conjecture relates the rank to the order of vanishing of $L(E_k, s)$ at $s = 1$, which motivates considering this feature even though our problem concerns the value at $s = 2$.
- 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.

The features factor(N_k) and factor(Δ_k) are encoded as lists of prime–exponent pairs, with the positive prime factors in increasing order. Each list begins with `N_Factor` or `D_Factor` and ends with `END_N_Factor` or `END_D_Factor`, respectively. Each pair begins with the token `PAIR`, followed by the prime and its exponent. Each integer is represented by a sign token (`INT+` or `INT-`), followed by the digits of its absolute value in the base used for the task, with one digit per token and the most significant digit first. Negative exponents in the factorization of Δ_k are retained, and an initial pair $(-1, 1)$ records a negative discriminant.

5. Machine learning results

The machine-learning experiments in this section investigate predictive relationships within the computed range $1 \le k \le 250{,}000$, with $k \neq 4$. Their purpose is to examine which arithmetic features are useful for prediction and how the model's predictions relate to the patterns found in our statistical analysis. In particular, the improved accuracy at the second 2-adic plateau suggests predictive relationships within this range that are not fully captured by $\hat{P}(k)$. This observation does not identify a general arithmetic rule or establish accuracy for substantially larger values of k.

For the valuation tasks at $p = 3, 5, 7$, we compare with the constant prediction given by the most frequent value of $v_p(n_k)$ in the training set. Table 5.1 reports its accuracy on the same 3,000 validation examples used in these experiments. For $p = 5, 7$, the constant prediction 0 has 100% recall on the zero class and 0% recall on every positive valuation class. The corresponding comparison for $p = 2$, including balanced accuracy, is reported on the separate test set in Table 5.2.

p	Constant prediction	Correct predictions	Accuracy
3	1	967/3,000	32.23%
5	0	2,210/3,000	73.67%
7	0	2,546/3,000	84.87%

Table 5.1. Majority-class baseline accuracies on the validation set for the valuation tasks at $p = 3, 5, 7$. Each constant prediction is determined from the training set.

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.

Including the root number ω_k as an input feature approximately doubles the factor-two accuracy of the greedy predictions, see Figure 5.1. For this comparison, we use three training runs for each choice of input features. In each run, we select the epoch with the highest validation factor-two accuracy among the first 300 epochs. The corresponding mean test accuracies are $(25.27 \pm 0.33)\%$ with k alone and $(49.40 \pm 1.25)\%$ with (k, ω_k), where the values after $\pm$ are sample standard deviations over the three runs. This improvement is consistent with the role of ω_k in determining the sign of r_k. We therefore include ω_k in most of the subsequent experiments.

In the context described above, providing the conductor N_k gives the fastest initial learning, while providing factor(Δ_k) gives faster initial learning than factor(N_k), see Figure 5.2. For each choice of input features, we use

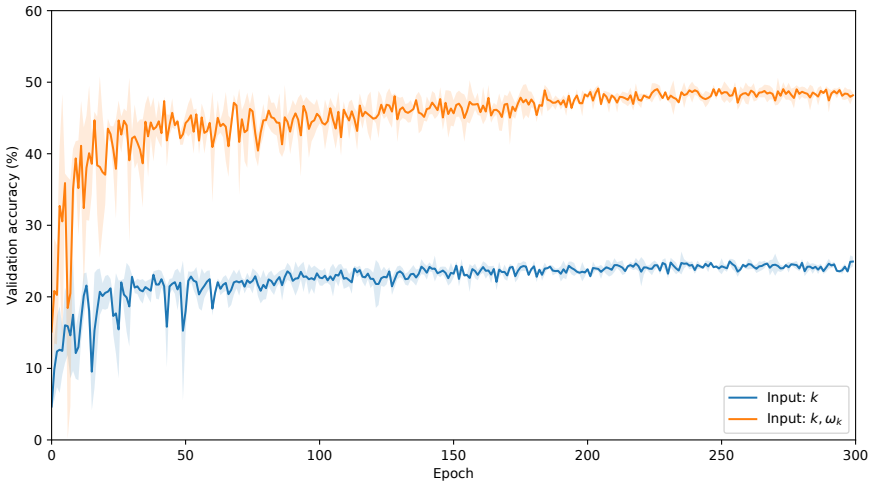


FIGURE 5.1. Validation factor-two accuracy for greedy predictions of r_k, using k alone or (k, ω_k) as input. For each choice of input features, the curve shows the mean over three runs and the shaded band shows the mean plus or minus one sample standard deviation. Log epochs $0, \ldots, 299$ are shown.

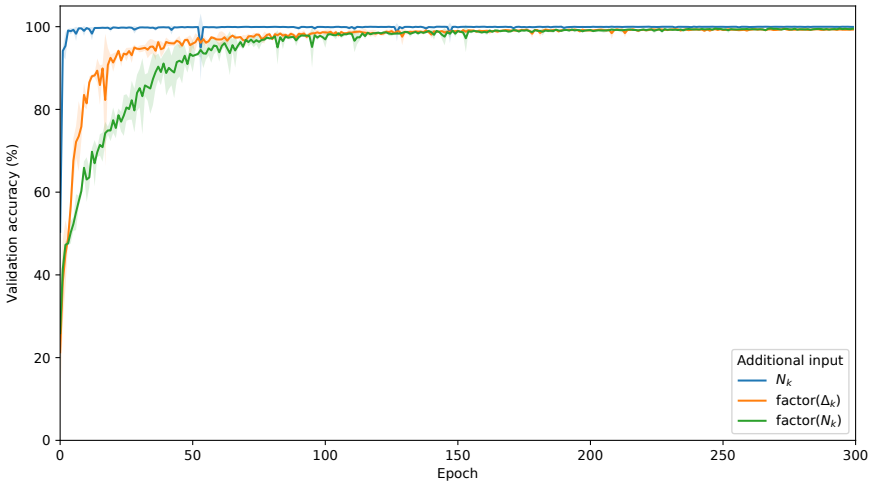


FIGURE 5.2. Validation factor-two accuracy for greedy predictions of r_k from (k, ω_k) with the additional feature N_k, factor(N_k), or factor(Δ_k). For each choice of input features, the curve shows the mean over three runs and the shaded band shows the mean plus or minus one sample standard deviation. Log epochs $0, \ldots, 299$ are shown.

three training runs. In each run, we select the earliest epoch attaining the highest validation factor-two accuracy among the first 300 epochs. The corresponding mean test accuracies are $(99.98 \pm 0.02)\%$ with N_k, $(99.42 \pm 0.14)\%$ with factor(N_k), and $(99.22 \pm 0.17)\%$ with factor(Δ_k). The high accuracy obtained when N_k is supplied is consistent with (3.4), which relates the scale of $|n_k|$ to $N_k/\log(k)$. Moreover, we speculate that the model recovers N_k from factor(Δ_k) and factor(N_k) and is unable to recover N_k from Δ_k (due to the fact that factorization is difficult). Nevertheless, we find it puzzling that factor(Δ_k) leads to faster learning than factor(N_k). The other considered features, namely the rank bounds and the a_p's, do not seem to influence the model's learning.

For comparison, replacing $L(E_k, 2)$ by 1 in (3.3) leads to the explicit arithmetic predictor

$$r_k^{\text{base}} := \omega_k \frac{4\pi^2 \mu_k}{N_k},$$

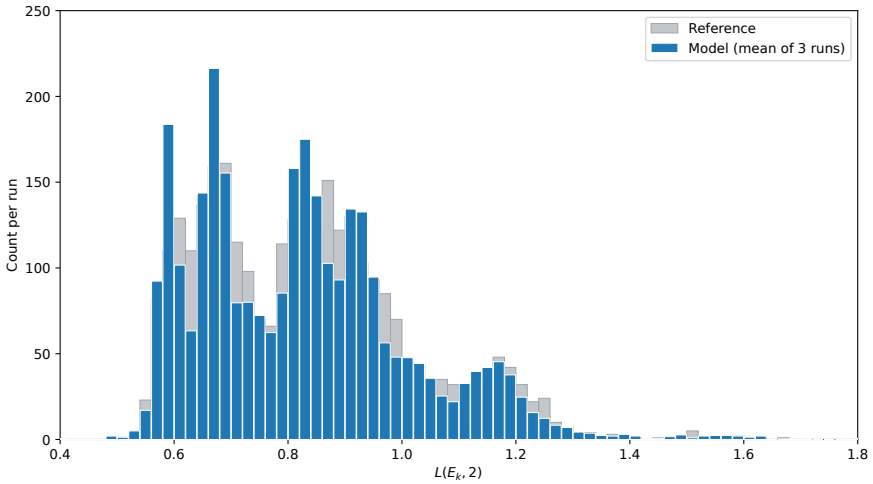


FIGURE 5.3. Distributions of the reference and model-induced values of $L(E_k, 2)$ on the 3,000 test examples, for inputs k, ω_k and factor(Δ_k). The grey histogram gives the reference values reconstructed from the stored rational candidates r_k. The blue histogram shows the mean bin counts over three runs, at validation-selected log epochs 273, 125, and 285 in run order. Both histograms use bins of width 0.02. The displayed interval $[0.4, 1.8]$ contains all reference values and more than 99% of the induced values in each run.

where μ_k is computed from k using (2.1). Thus this predictor uses the inputs k, ω_k and N_k. Equations (0.1) and (3.3) give

$$\frac{r_k^{\text{base}}}{r_k} = L(E_k, 2).$$

The observed range $(0.5222, 1.7532)$ for $L(E_k, 2)$, reported in Section 3, is contained in $(1/2, 2)$. This implies that the arithmetic predictor has the correct sign and satisfies the factor-two criterion throughout the main dataset $1 \le k \le 250{,}000$, $k \ne 4$. Thus this arithmetic baseline also achieves 100% factor-two accuracy on these data. The neural-network experiments show how the representation of the input features affects the learning of this scale.

Finally, we use (0.1) and (3.3) to associate to each nonzero greedy prediction s_k the value

$$\omega_k \frac{4\pi^2 \mu_k}{N_k s_k}.$$

Figure 5.3 compares the distribution of these induced values of $L(E_k, 2)$ with the reference values reconstructed from the stored rational candidates r_k, on the same 3,000 test examples. We use the three runs with inputs k, ω_k and factor(Δ_k), each at its validation-selected epoch. The central part of the induced distribution resembles the reference distribution, although a small number of induced values lie outside the observed reference range. This comparison concerns marginal distributions and does not by itself establish accurate pointwise prediction of $L(E_k, 2)$.

5.2. **p-adic valuations of n_k for $p \ge 5$.** These experiments were performed for $p = 5, 7$. They had $v_p(n_k)$ as output, with k as the sole input feature or with the additional feature factor(N_k); for $p = 5$, we also used factor(Δ_k). All inputs and outputs were written in base p. The constant prediction $v_p(n_k) = 0$ achieves validation accuracies of 73.67% for $p = 5$ and 84.87% for $p = 7$. Under the reference distribution in Section 3.2, the corresponding probability is $(p-2)/(p-1)$. Two runs for $p = 5$ using factor(N_k) correctly predict some positive valuations, but their overall validation accuracy at the final epoch is below that of the constant-zero predictor.

5.3. **3-adic valuations of n_k.** For $p = 3$, we consider three training runs with k and factor(Δ_k) as input features, with all integers encoded in base 3. We express the targets and predictions in terms of $v_3(n_k) = -v_3(r_k)$. Figure 5.4 shows the validation greedy exact-match accuracy over the course of training. For each run, we select the earliest epoch attaining the highest validation accuracy among the first 300 epochs. The selected epochs are 244, 216, and 298. On the separate set of 3,000 test examples, the corresponding individual exact-match accuracies are 49.20%, 48.40%, and 48.80%, respectively. The mean accuracy is (48.80 ± 0.40)%, where the value after $\pm$ is the sample

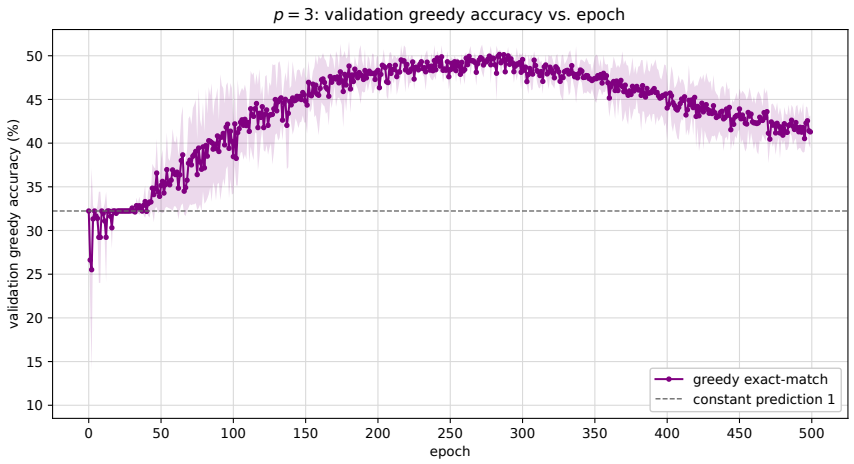


FIGURE 5.4. Validation greedy exact-match accuracy for $v_3(n_k)$, with inputs k and factor(Δ_k) encoded in base 3. The curve shows the mean over three runs, and the shaded band shows one sample standard deviation on either side of the mean. The dashed line shows the accuracy of the constant prediction 1 from Table 5.1. All 500 completed training epochs are shown; epoch selection is restricted to the first 300.

standard deviation over the three runs. On the same test set, the constant prediction $v_3(n_k) = 1$ has an accuracy of 31.17%. The validation accuracy declines during the later part of training.

Figure 5.5 shows the validation greedy exact-match accuracy conditioned on the true valuation $v_3(n_k) = 0$, $v_3(n_k) = 1$, or $v_3(n_k) > 1$. At many early epochs, the model predicts $v_3(n_k) = 1$ for every validation example, which explains the initial plateau in Figure 5.4. The improvement over this constant prediction comes mainly from learning to recognize the cases with $v_3(n_k) = 0$. At the validation-selected epochs used above, the mean conditional exact-match accuracies on the test set are $(68.71 \pm 1.23)\%$ for $v_3(n_k) = 0$, $(95.40 \pm 0.81)\%$ for $v_3(n_k) = 1$, and $(1.77 \pm 0.92)\%$ for $v_3(n_k) > 1$. Thus, exact prediction of the higher valuations remains poor. During the later part of training, the validation accuracy on the $v_3(n_k) = 1$ cases declines, contributing to the decrease in overall accuracy seen in Figure 5.4.

5.3.1. *Learning on different congruence classes.* In our statistical analysis, we observed that $v_3(n_k) \geq 1$ for $k \equiv 0, \pm 5 \bmod 27$, apart from a few exceptions in the dataset. At the validation-selected epochs used above, the model predicts $v_3(n_k) = 1$ with a mean frequency of $(97.27 \pm 2.44)\%$ on the 354 test examples in these classes. Its mean exact-match accuracy on these examples is $(41.90 \pm 0.16)\%$, compared with 42.09% for the constant prediction 1 and $(48.80 \pm 0.40)\%$ on the full test set. Thus, at these selected epochs, accuracy on these congruence classes is very close to that of the constant predictor.

5.3.2. *Learning on the sets from Table 3.5.* Next we examine the model's predictions on the set A, for which the empirical parity laws (3.7) were discussed above. Figure 5.6 shows the validation greedy exact-match accuracy for $k \in A$, conditioned on the true valuation $v_3(n_k) = 0$, $v_3(n_k) = 1$, or $v_3(n_k) > 1$.

At the validation-selected epochs used above, the mean conditional test accuracies on A are $(96.60 \pm 0.37)\%$ for $v_3(n_k) = 0$ and $(99.78 \pm 0.39)\%$ for $v_3(n_k) = 1$. On A, the model therefore distinguishes the cases with valuation 0 from those with valuation 1 with high accuracy. For the 180 test examples in A with $v_3(n_k) > 1$, the mean exact-match accuracy at these selected epochs is $(0.19 \pm 0.32)\%$. Figure 5.7 compares greedy exact-match accuracy on the sets in Table 3.5, restricted to examples with true $v_3(n_k) \in \{0, 1\}$. The table reports test results at the validation-selected epochs used above, and the curves show mean validation accuracy over the three runs.

In this comparison, the highest mean test accuracy is attained on the set A. Although Table 3.5 records no exceptions to the parity laws 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. Excluding the values $k \equiv 4 \bmod 8$, passing from B to B' and from C to C', increases the mean conditional test accuracy.

Finally, we consider the values of k outside $C'' \cup \{k \equiv 0, \pm 5 \bmod 27\}$, restricting to those for which $v_3(n_k) \in \{0, 1\}$. Figure 5.8 shows the mean validation accuracy on these examples. After an initial increase, the accuracy declines in

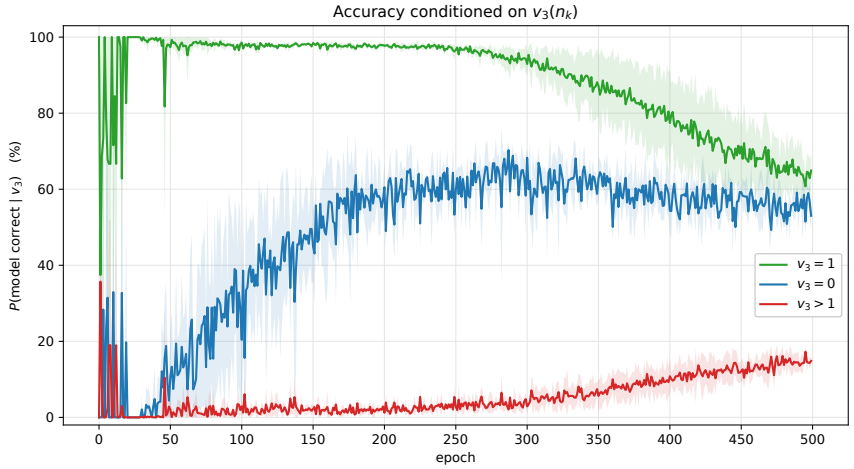


FIGURE 5.5. Validation greedy exact-match accuracy conditioned on the true valuation $v_3(n_k)$. The green, blue and red curves correspond to $v_3(n_k) = 1$, $v_3(n_k) = 0$ and $v_3(n_k) > 1$, respectively. Each curve shows the mean over three runs; the shaded bands show one sample standard deviation, restricted to $[0, 100]\%$ for display. In the group $v_3(n_k) > 1$, a prediction is counted as correct only if it gives the exact valuation. All 500 completed training epochs are shown.

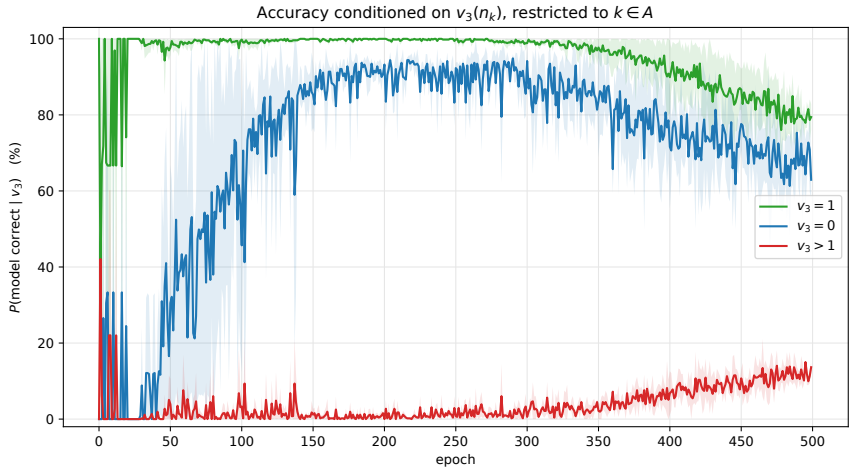


FIGURE 5.6. Validation greedy exact-match accuracy on $k \in A$, conditioned on the true valuation $v_3(n_k)$. The green, blue and red curves correspond to $v_3(n_k) = 1$, $v_3(n_k) = 0$ and $v_3(n_k) > 1$, respectively. Each curve shows the mean over three runs; the shaded bands show one sample standard deviation, restricted to $[0, 100]\%$ for display. For $v_3(n_k) > 1$, correctness still requires the exact valuation. All 500 completed training epochs are shown.

the later stages of training. At the validation-selected epochs used above, the mean test accuracy on the 551 such examples is $(65.34 \pm 0.96)\%$. On the same test examples, the constant predictor $v_3(n_k) = 1$ has accuracy 49.91%. Thus the model also achieves accuracy above this baseline on the remaining examples within the computed range.

Using the same rule for selecting epochs from validation accuracy in each configuration, the three runs with inputs k and factor(N_k) give a mean greedy exact-match accuracy of $(47.06 \pm 0.30)\%$ on the full test set of 3,000 examples, compared with $(48.80 \pm 0.40)\%$ for inputs k and factor(Δ_k). With the conductor factorization, mean validation accuracy rises earlier in training: at epoch 50 it is 43.74%, compared with 35.16% for the discriminant factorization. Using k, factor(N_k), and factor(Δ_k) together as inputs gives a mean greedy exact-match test accuracy of $(47.49 \pm 0.59)\%$ under the same validation-based epoch-selection rule. At epoch 50, mean validation accuracy

set	n	acc. (%)
A	306	98.15 ± 0.38
A'	340	95.88 ± 1.02
A''	566	92.82 ± 1.30
B	1,133	86.82 ± 0.59
B'	977	91.98 ± 0.97
B''	937	91.89 ± 1.08
C	1,251	85.74 ± 0.60
C'	1,081	90.63 ± 0.94
C''	1,035	90.43 ± 1.05

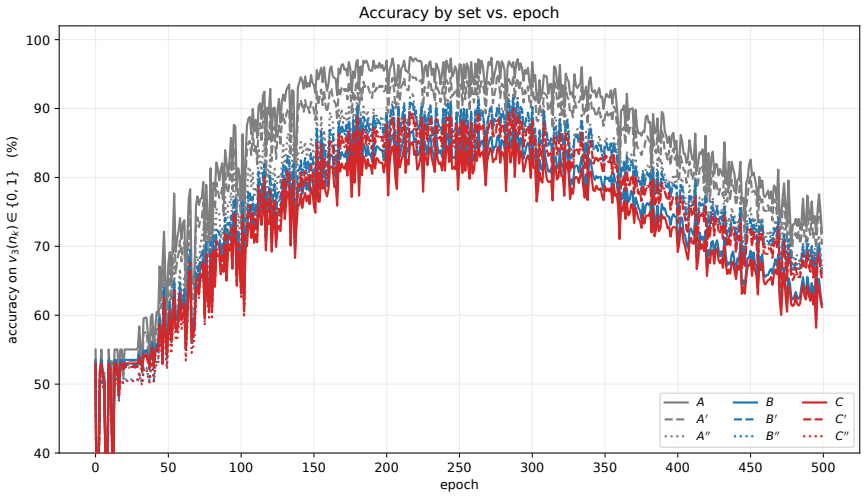


FIGURE 5.7. Greedy exact-match accuracy restricted to true $v_3(n_k) \in \{0, 1\}$ on the sets in Table 3.5. Left: mean test accuracy, with sample standard deviation over three runs, at the validation-selected epochs used above; n is the number of test examples in each set with true $v_3(n_k) \in \{0, 1\}$. Right: mean validation accuracy over the same three runs, as a function of the training-log epoch, numbered from 0. The vertical axis starts at 40%.

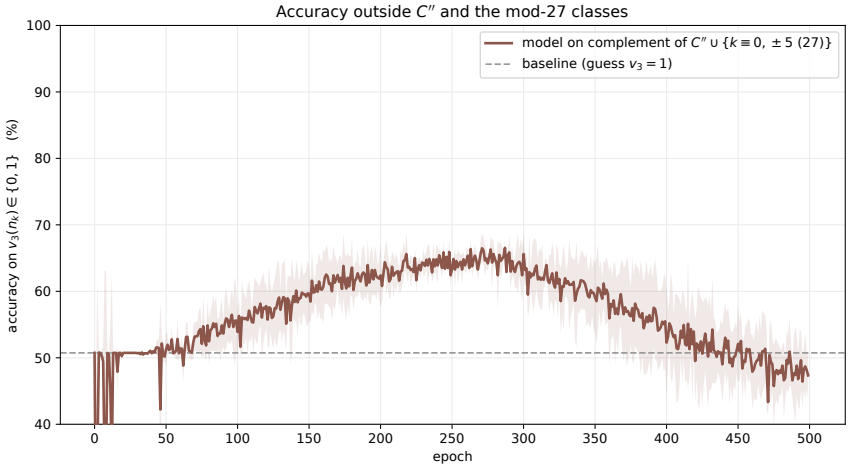


FIGURE 5.8. Validation accuracy outside $C'' \cup \{k \equiv 0, \pm 5 \bmod 27\}$, restricted to the 540 validation examples with true valuation $v_3(n_k) \in \{0, 1\}$. The brown curve is the mean greedy exact accuracy over three runs, and the shaded band shows one sample standard deviation on either side of the mean. The dashed grey line is the accuracy of the constant predictor $v_3(n_k) = 1$ on the same validation examples, namely 50.74%. The vertical axis starts at 40%.

for this configuration is 37.74%, compared with 43.74% for inputs k and factor(N_k). By contrast, supplying N_k or Δ_k themselves, rather than their factorizations, does not lead to appreciable learning in these experiments. In both configurations, the validation accuracy remains essentially at the majority-class baseline obtained by always predicting $v_3(n_k) = 1$. At the validation-selected checkpoints, the mean test accuracies are $(31.17 \pm 0.00)\%$ with N_k and $(31.21 \pm 0.08)\%$ with Δ_k, compared with 31.17% for the constant prediction $v_3(n_k) = 1$ on the same test set. This is consistent with the factorization of these quantities, rather than merely their size, being the useful arithmetic input for this task.

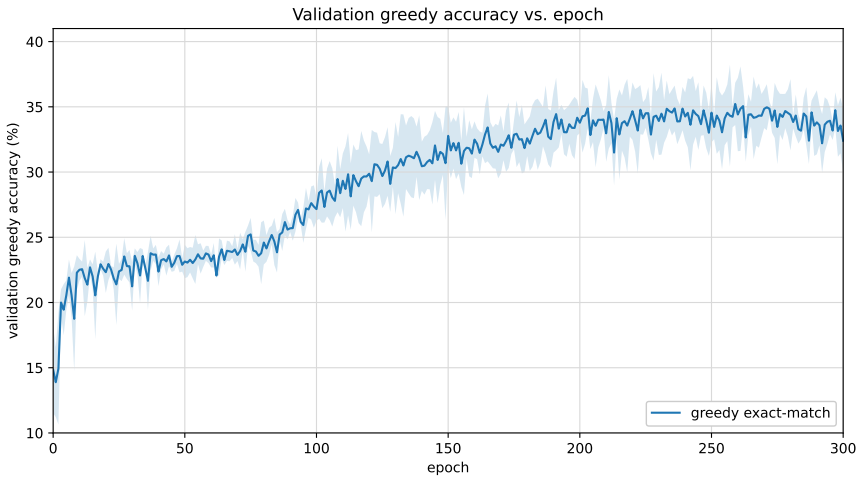


FIGURE 5.9. Validation greedy exact-match accuracy for the 2-adic experiment with inputs k and factor(Δ_k). The curve is the mean over three runs, and the shaded region shows one sample standard deviation on either side of the mean.

5.4. **2-adic valuations of n_k.** For $p = 2$, we consider three runs in which the model is given k and the factorization of the discriminant factor(Δ_k) as input and $v_2(n_k)$ as output, all written in base 2. Figure 5.9 shows the mean validation accuracy over these runs. The model quickly reaches an accuracy of about $23\% - 25\%$, giving a first plateau during the early stages of training. The accuracy then begins to rise around epochs 80–100 and reaches a second plateau around $34\% - 35\%$. Using the same validation-based epoch-selection rule as above, the selected epochs for the three runs are 261, 272, and 211. The corresponding mean accuracy on the full test set of 3,000 examples is $(37.12 \pm 1.60)\%$, where the value after $\pm$ is the sample standard deviation over the three runs. We analyze the behavior of the model at the two plateaus below.

For each of the three runs, we select the epoch with the highest validation greedy exact-match accuracy among the first 300 epochs. The selected epochs are 261, 272, and 211. We report the archived predictions for these checkpoints on the separate test set of 3,000 examples, which is not used for epoch selection. The corresponding test accuracies are 37.63%, 35.33%, and 38.40%, giving a mean accuracy of $(37.12 \pm 1.60)\%$.

On the same test set, the predictor $\hat{P}(k)$ is correct on 737 examples, giving an accuracy of 24.57%. As a majority-class baseline, we use the constant prediction 10, the most frequent value of $v_2(n_k)$ in the training set. It is correct on 392 of the 3,000 test examples, giving an accuracy of 13.07%.

We also report balanced accuracy: for each true valuation present in the test set, we compute the proportion of correct predictions and then take the unweighted average over these classes. The 23 classes are $v_2(n_k) = 2, \ldots, 24$. The comparison is given in Table 5.2.

Predictor	Exact-match accuracy	Balanced accuracy
Model (three runs)	$(37.12 \pm 1.60)\%$	$(27.40 \pm 0.32)\%$
$\hat{P}(k)$	24.57%	13.08%
Constant prediction 10	13.07%	4.35%

TABLE 5.2. Exact-match and balanced accuracies on the separate test set of 3,000 examples. For the model, we report the mean and sample standard deviation over three runs, with the checkpoint in each run selected using validation accuracy.

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. Figure 5.10 shows the mean proportion of validation examples for which the model's predicted valuation is at least $\hat{v}(k)$. The proportion rises very quickly: by epoch 7 its mean value is

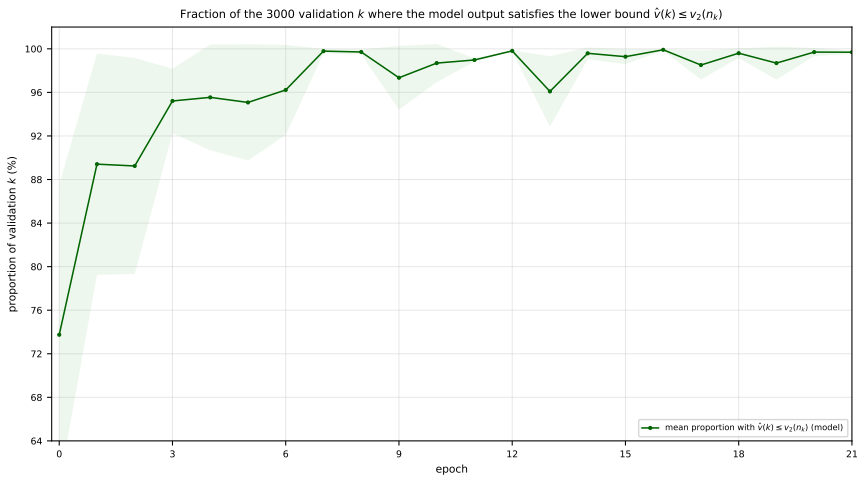


FIGURE 5.10. The proportion of k in the validation set for which the model's predicted valuation is at least $\hat{v}(k)$. The green curve is the mean over three runs, and the shaded region shows one sample standard deviation on either side of the mean.

99.79%, and from that point through epoch 21 it remains above 96% and is usually close to 100%. At the validation-selected checkpoints used above, the corresponding mean proportion on the separate test set is $(99.86 \pm 0.08)\%$.

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 predictions on the validation set by computing their Pearson correlation coefficient at each epoch; see Figure 5.12. The mean correlation over the three runs rises rapidly and reaches its maximum, $r \approx 0.971$, at epoch 62. It then decreases overall, reaching $r \approx 0.876$ at epoch 300. At the validation-selected checkpoints used above, the mean correlation is 0.918 ± 0.008.

The increase in validation accuracy after the first plateau, together with the decrease in correlation with $\hat{P}(k)$, suggests that the model improves its predictions while gradually departing from this explicit predictor. These observations do not identify the arithmetic features responsible for the improvement. 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 $r = 0.78$. Figure 5.11 gives another view of the same phenomenon.

For each displayed epoch, we pool the validation predictions from the three runs. The selected-checkpoint panel pools the validation predictions from each run's own validation-selected checkpoint. The upper subfigure compares the model's predictions with $\hat{P}(k)$, while the lower subfigure compares them with the true valuations $v_2(n_k)$.

The epoch 40 panel for $\hat{P}(k)$ shows a strong concentration near the diagonal during the first plateau. The pooled Pearson correlation with $\hat{P}(k)$ is 0.94 at epoch 100 and 0.87 at epoch 316. The corresponding correlations with the true valuations are 0.74 and 0.71, respectively. Thus the increase in accuracy beyond the first plateau is not explained simply by the model becoming a closer approximation to $\hat{P}(k)$.

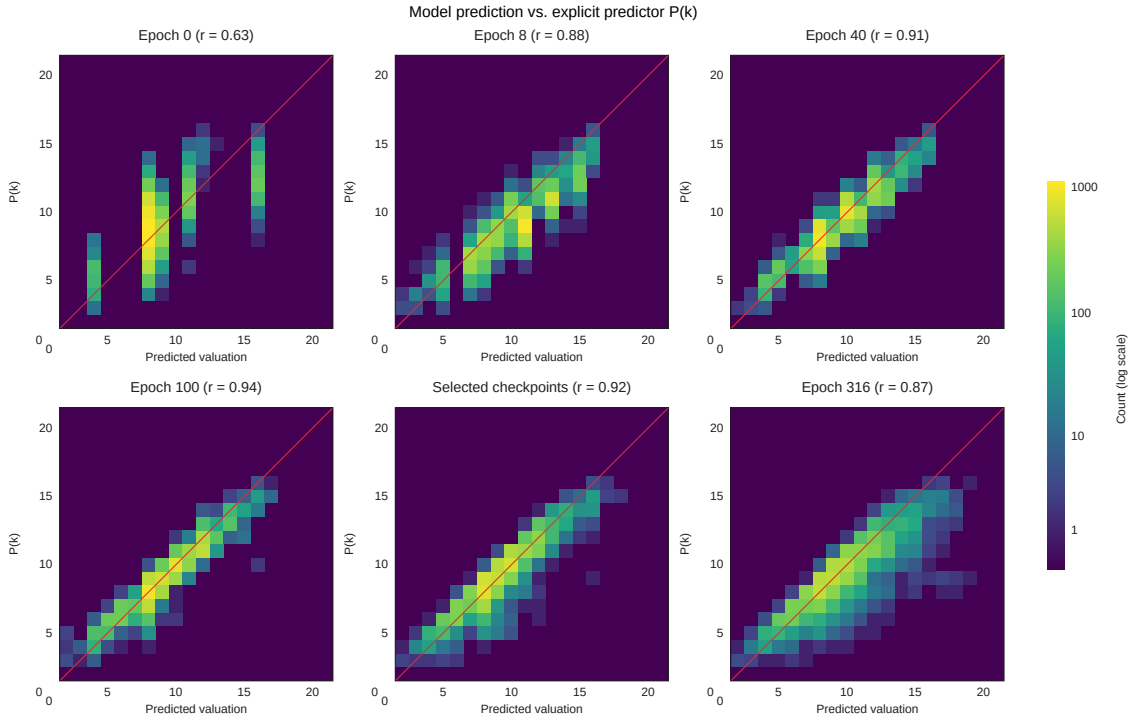
On the validation set, the predictor $\hat{P}(k)$ is correct for 719 of the 3,000 examples, giving an accuracy of 23.97%. The model's accuracy on the first plateau is approximately 23%–25%. Together with the confusion matrices, these similar accuracies support the interpretation that the model's predictions approximate $\hat{P}(k)$ during the first plateau.

5.4.2. *Learning around the second plateau of Figure 5.9.* The plateau diagnostics below use the validation predictions from the three runs. For the second plateau, it becomes much less clear *what* the model actually learned. To investigate this, we consider the proportions

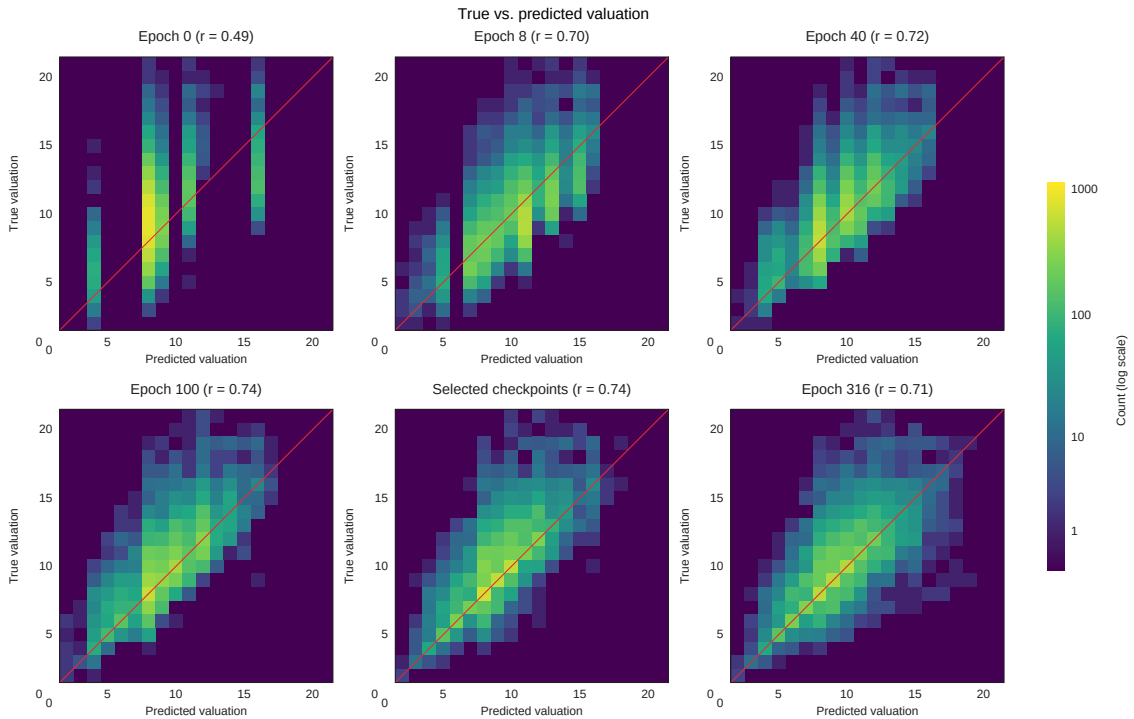
$$q_g := \mathbb{P}\big(\text{model is correct} \mid G(k) = v_2(n_k) - \hat{v}(k) = g\big)$$

on the validation set. When $g = 0$, the lower bound in Conjecture 3.2 is sharp. The mean q_0 across the three runs increases from 23.13% at epoch 40 to 61.67% at the validation-selected checkpoints (epochs 261, 272, and 211, in run order). Figure 5.13 shows how q_0 changes over training.

From epoch 40 to the validation-selected checkpoints, the mean q_1 increases from 37.37% to 68.85%, while the mean q_2 decreases from 30.73% to 24.50%. The largest contributions to the increase in overall validation accuracy



(A) Model prediction vs. predictor $\hat{P}(k) = \mathrm{round}(\hat{v}(k) + s(k) + 1)$.



(B) Model prediction vs. true $v_2(n_k)$.

FIGURE 5.11. Confusion matrices at epochs 0, 8, 40, 100, and 316, and at the validation-selected checkpoints (epochs 261, 272, and 211 in run order). Each panel pools up to 9,000 valid predictions from three runs on the same 3,000 validation examples. The color scales show counts on a logarithmic scale. Panel titles give the pooled Pearson correlation coefficient r.

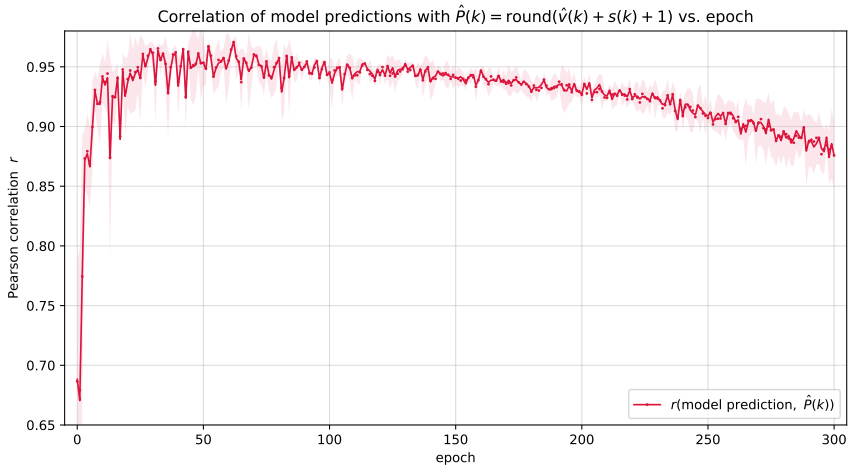


FIGURE 5.12. The Pearson correlation between the model's greedy predictions and our predictor $\hat{P}(k)$ on the validation set as a function of the epoch. The curve is the mean over three runs, and the shaded region shows one sample standard deviation on either side of the mean.

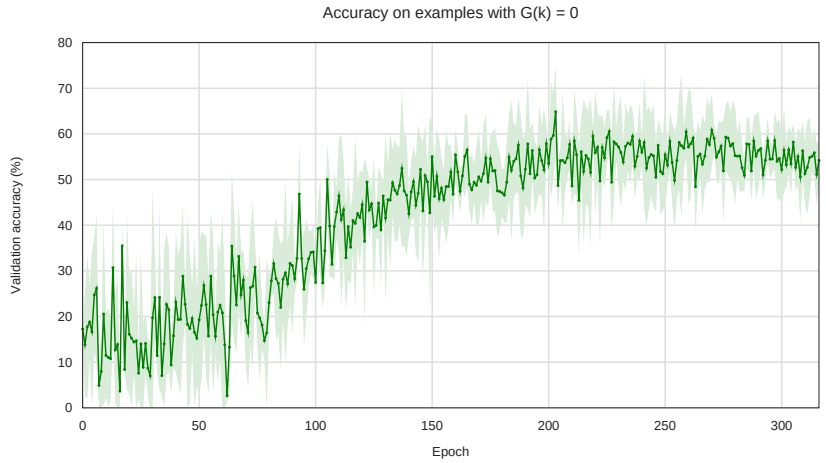


FIGURE 5.13. The percentage of correct greedy predictions among the 647 validation examples satisfying $v_2(n_k) = \hat{v}(k)$, as a function of the epoch. The curve shows the mean across the three runs, and the shaded region shows one sample standard deviation on either side of the mean. Epochs 0 through 316 are shown.

come from $g = 0$ and $g = 1$, with about 249 and 226 additional correct predictions per run on average, respectively. See Table 5.3.

To test whether the improvement over $\hat{P}(k)$ persists on contiguous held-out intervals, we repeated the 2-adic experiment with inputs k and factor(Δ_k) encoded in base 2, using ten random seeds for each of two splits. In the first split, we trained on $1 \le k \le 244{,}000$ $(k \ne 4)$, used $244{,}001 \le k \le 247{,}000$ for validation, and held out $247{,}001 \le k \le 250{,}000$ for testing. In the second split, we trained on $[1, 125{,}000] \cup [131{,}001, 250{,}000]$, excluding $k = 4$, used $125{,}001 \le k \le 128{,}000$ for validation, and held out $128{,}001 \le k \le 131{,}000$ for testing. In each run, we selected the earliest epoch attaining the highest validation greedy exact-match accuracy among epochs $0, \dots, 299$, and report the corresponding archived test predictions. On the upper test block, mean exact-match accuracy rose from 23.38% at epoch 40 to $(37.61 \pm 0.91)\%$ at the selected checkpoints, compared with 24.83% for $\hat{P}(k)$. On the interior test block, it rose from 23.45% to $(37.86 \pm 1.34)\%$, compared with 24.03% for $\hat{P}(k)$. All comparisons within a

$G(k)$	Count	Epoch 40 (%)	Selected checkpoint (%)
0	647	23.13 ± 15.38	61.67 ± 11.49
1	718	37.37 ± 16.85	68.85 ± 2.37
2	679	30.73 ± 22.82	24.50 ± 4.27
3	411	10.46 ± 8.43	4.79 ± 2.44
4	268	0.50 ± 0.43	2.49 ± 0.78
≥ 5	277	0.00 ± 0.00	1.20 ± 0.42

TABLE 5.3. Exact-match validation accuracy by excess $G(k)$. Both accuracy columns use the same 3,000 validation examples. Values are means and sample standard deviations across three runs. The validation-selected checkpoints are at epochs 261, 272, and 211, in run order.

split use the same 3,000 test examples. At the selected checkpoints, every run of both splits has higher exact-match accuracy than $\hat{P}(k)$. Thus the improvement persists on both contiguous held-out intervals.

The learning observed with factor(Δ_k) as a feature can be interpreted in light of equations (3.9) and (3.11). The prime factors and their exponents in the discriminant provide information about the contributions from k and $k^2 - 16$, whose odd prime divisors have different weights in $\hat{v}(k)$. Among the archived runs, all of which include k as an input, the highest greedy validation accuracies over all 500 epochs are 37.63% with factor(Δ_k), 26.87% with factor(N_k), and 36.47% with both factorizations. The highest value with factor(Δ_k) and c_2 is 38.00%. Across runs with input pairs (k, c_2), (k, N_k), and (k, Δ_k), peak validation accuracies range from 16.03% to 17.27%.

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 difference between the behavior at different primes. For primes $p \geq 5$, the observed positive valuation frequencies are close to p^{-r} for $r \geq 1$. At $p = 5$, the agreement improves substantially when we restrict to curves with good reduction at 5.

The primes 2 and 3 exhibit richer arithmetic. At 3, congruence conditions on k interact with the parity of the number of bad primes congruent to 1 modulo 3, producing observed parity constraints on large subfamilies. At 2, a weighted count of the odd primes dividing k and $k^2 - 16$, together with a correction arising from the reduction at 2, leads to the lower bound in Conjecture 3.2. The empirical mean of the excess over this bound depends on the 2-adic class of k and is approximately described by $2 + s(k)$. For the overall valuation histogram, we use the reference model in (3.11), which combines $\hat{v}(k) + \mathrm{round}(s(k))$ with an auxiliary negative-binomial variable. This gives a first-order description of the observed distribution.

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. For $p = 5$ and $p = 7$, the frequent zero class provides a strong baseline, illustrating the difficulty caused by highly unbalanced valuation distributions. Some runs at $p = 5$ correctly predict positive valuations, but their overall accuracy at the final epoch is below that of the constant-zero predictor.

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. It also exceeds the constant-prediction baseline on the tested examples with $v_3(n_k) \in \{0, 1\}$ outside $C'' \cup \{k \equiv 0, \pm 5 \bmod 27\}$. 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 $\widehat{P}(k)$ derived from our statistical model. During the second plateau, however, its accuracy increases while its correlation with $\widehat{P}(k)$ decreases. Across the three 2-adic runs, the largest contributions to the increase in validation accuracy from epoch 40 to the validation-selected checkpoints come from examples with $G(k) = 0$ and $G(k) = 1$. The three validation-selected checkpoints achieve a mean test

exact-match accuracy of $(37.12 \pm 1.60)\%$, compared with 24.57% for $\widehat{P}(k)$ on the same 3,000 test examples. This behavior suggests predictive relationships within the computed range that are not fully captured by the explicit predictor $\widehat{P}(k)$. The improvement also persists in the two blocked-split experiments. One tests prediction beyond the largest training value, while the other tests a contiguous omitted interval with training values on both sides. These results do not establish accuracy for substantially larger values of k or identify a general arithmetic rule.

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 understand the distribution of the excess $G(k)$ and its dependence on the 2-adic class of k. 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 reference model assigns probability p^{-r} to valuation $r \geq 1$. Hence, in this model, the probability of a positive valuation is

$$\sum_{r=1}^{\infty} p^{-r} = \frac{1}{p-1}.$$

Conditional on a positive valuation, the distribution is the usual geometric distribution. The main arithmetic question concerns this increased probability of the first factor of p, compared with $1/p$ in a uniform-distribution model modulo p.

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 $G(k) = v_2(n_k) - \widehat{v}(k)$, 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 $p = 5$ and $p = 7$, $v_p(n_k) = 0$ is a strong baseline. This constant prediction achieves high overall accuracy because zero is the most frequent observed class, while failing to identify positive valuations. 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 improve prediction of the rarer positive valuations. A first natural extension, taking $k = \sqrt{m}$ for $m \in \mathbb{Z}$ of either sign, is examined in Section 3.5. In the computed range, the numerically recognized rational candidates $r_{\sqrt{m}}$ again have near-integral reciprocals. An analogue of the 2-adic lower bound persists, and the odd-prime frequencies remain broadly similar. For $p = 5$, however, the excess of positive valuations persists after restricting to good reduction. 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.

(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.

k	$k-4$	$k+4$	$W(k)$	$\hat{v}(k)$	$v_2(n_k)$	$v_2(n_k)-\hat{v}(k)$
321559	$3\cdot5\cdot13\cdot17\cdot97$	$11\cdot23\cdot31\cdot41$	21	17	19	+2
400201	$3\cdot7\cdot17\cdot19\cdot59$	$5\cdot13\cdot47\cdot131$	21	17	20	+3
440891	$23\cdot29\cdot661$	$3\cdot5\cdot7\cdot13\cdot17\cdot19$	21	17	19	+2
471709	$3\cdot5\cdot13\cdot41\cdot59$	$11\cdot19\cdot37\cdot61$	21	17	21	+4
474415	$3\cdot7\cdot19\cdot29\cdot41$	$11\cdot17\cdot43\cdot59$	21	17	18	+1
519589	$3\cdot5\cdot11\cdot47\cdot67$	$19\cdot23\cdot29\cdot41$	21	17	20	+3
572029	$3\cdot5^2\cdot29\cdot263$	$7\cdot11\cdot17\cdot19\cdot23$	21	17	22	+5
729491	$11\cdot17\cdot47\cdot83$	$3^2\cdot5\cdot13\cdot29\cdot43$	21	17	21	+4
889291	$3\cdot7\cdot17\cdot47\cdot53$	$5\cdot11\cdot19\cdot23\cdot37$	23	19	21	+2
1008011	$7\cdot11\cdot13\cdot19\cdot53$	$3\cdot5\cdot17\cdot59\cdot67$	22	18	20	+2
1394701	$3\cdot17\cdot23\cdot29\cdot41$	$5\cdot13\cdot43\cdot499$	22	18	21	+3
1601149	$3^2\cdot5\cdot7\cdot13\cdot17\cdot23$	$83\cdot101\cdot191$	22	18	21	+3
1847751	$11\cdot17\cdot41\cdot241$	$5\cdot7\cdot13\cdot31\cdot131$	22	18	23	+5
1944939	$5\cdot19\cdot59\cdot347$	$7\cdot11\cdot13\cdot29\cdot67$	22	18	22	+4

TABLE 7.6. Prime-rich odd values of k in $2.5\times10^5 < k < 2\times10^6$ chosen to maximise $W(k) := \omega_{\mathrm{odd}}(k) + 2\,\omega_{\mathrm{odd}}(k^2-16)$; 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 $v_2(n_k) \geq \hat{v}(k)$.

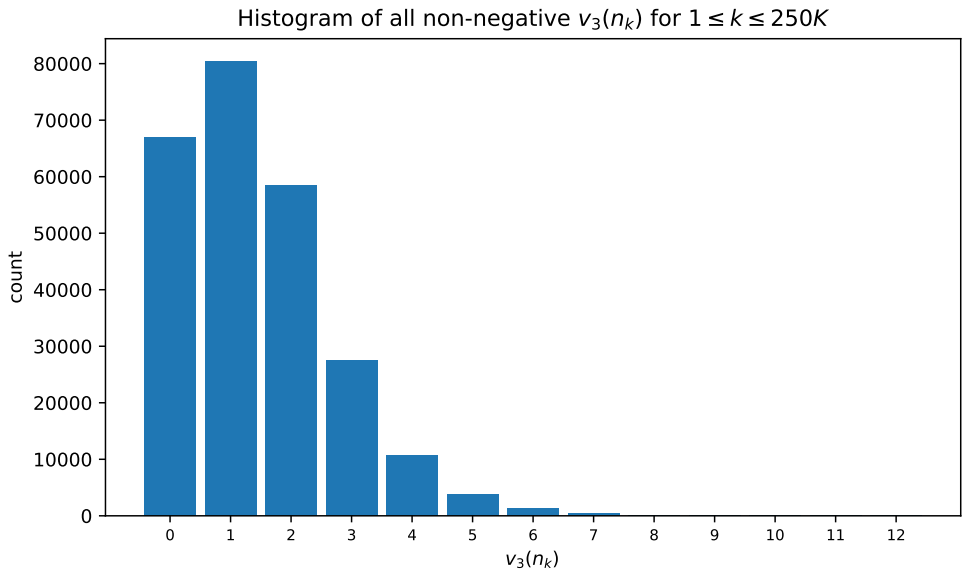


FIGURE 7.1. Histogram of the 3-adic valuations $v_3(n_k)$ for $1 \le k \le 250K$ such that $v_3(n_k) \ge 0$, where $K = 1000$.

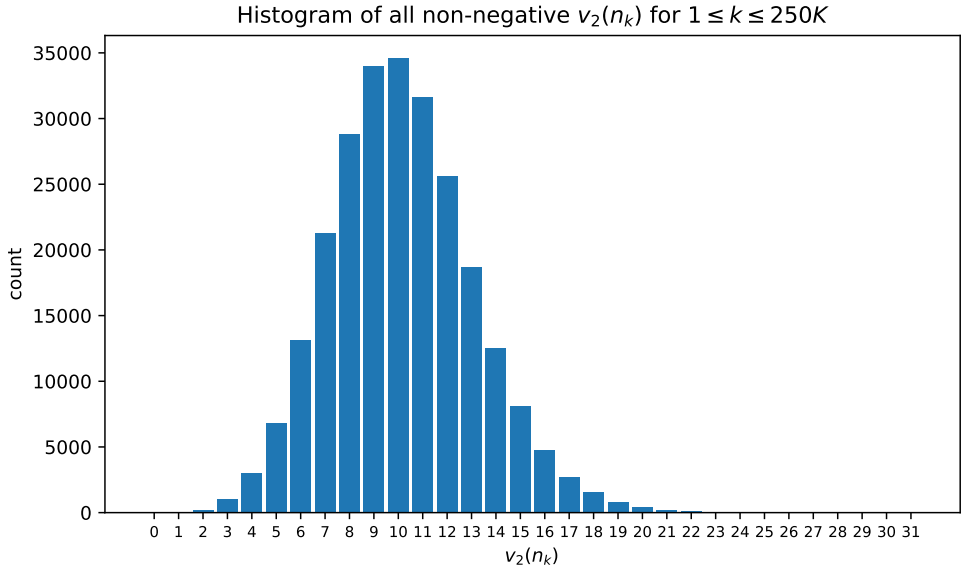


FIGURE 7.2. Histogram of the 2-adic valuations $v_2(n_k)$ for $1 \le k \le 250K$ such that $v_2(n_k) \ge 0$, where $K = 1000$.

k	$k-4$	$k+4$	$W(k)$	$\hat{v}(k)$	$v_2(n_k)$	$v_2(n_k)-\hat{v}(k)$
940474	$2\cdot3\cdot5\cdot23\cdot29\cdot47$	$2\cdot7\cdot11\cdot31\cdot197$	21	19	21	+2
1069814	$2\cdot5\cdot7\cdot17\cdot29\cdot31$	$2\cdot3\cdot37\cdot61\cdot79$	21	19	27	+8
1195674	$2\cdot5\cdot7\cdot19\cdot29\cdot31$	$2\cdot11\cdot17\cdot23\cdot139$	21	19	20	+1
1204606	$2\cdot3\cdot7\cdot23\cdot29\cdot43$	$2\cdot5\cdot11\cdot47\cdot233$	21	19	22	+3
1234874	$2\cdot5\cdot7\cdot13\cdot23\cdot59$	$2\cdot3\cdot29\cdot47\cdot151$	21	19	22	+3
1258786	$2\cdot3\cdot7\cdot17\cdot41\cdot43$	$2\cdot5\cdot13\cdot23\cdot421$	21	19	21	+2
1286666	$2\cdot13\cdot17\cdot41\cdot71$	$2\cdot3\cdot5\cdot7\cdot11\cdot557$	21	19	21	+2
1540136	$2^2\cdot11\cdot17\cdot29\cdot71$	$2^2\cdot3\cdot5\cdot7\cdot19\cdot193$	21	19	19	+0
1587226	$2\cdot3^3\cdot7\cdot13\cdot17\cdot19$	$2\cdot5\cdot23\cdot67\cdot103$	21	19	20	+1
1614826	$2\cdot3\cdot11\cdot43\cdot569$	$2\cdot5\cdot7\cdot17\cdot23\cdot59$	21	19	20	+1
1653526	$2\cdot3\cdot13\cdot17\cdot29\cdot43$	$2\cdot5\cdot37\cdot41\cdot109$	21	19	22	+3
1669906	$2\cdot3\cdot13\cdot79\cdot271$	$2\cdot5\cdot11\cdot17\cdot19\cdot47$	21	19	20	+1
1700846	$2\cdot11\cdot13\cdot19\cdot313$	$2\cdot3\cdot5^2\cdot17\cdot23\cdot29$	21	19	21	+2
1703464	$2^2\cdot3\cdot5\cdot11\cdot29\cdot89$	$2^2\cdot13\cdot17\cdot41\cdot47$	21	19	26	+7
1717166	$2\cdot41\cdot43\cdot487$	$2\cdot3\cdot5\cdot7\cdot13\cdot17\cdot37$	21	19	24	+5
1722886	$2\cdot3\cdot7\cdot17\cdot19\cdot127$	$2\cdot5\cdot13\cdot29\cdot457$	21	19	21	+2
1762094	$2\cdot5\cdot11\cdot83\cdot193$	$2\cdot3\cdot13\cdot19\cdot29\cdot41$	21	19	24	+5
1774630	$2\cdot3\cdot7\cdot29\cdot31\cdot47$	$2\cdot23\cdot173\cdot223$	21	19	22	+3
1791566	$2\cdot17\cdot23\cdot29\cdot79$	$2\cdot3\cdot5\cdot11\cdot61\cdot89$	21	19	24	+5
1812794	$2\cdot5\cdot7\cdot19\cdot29\cdot47$	$2\cdot3^2\cdot13\cdot61\cdot127$	21	19	19	+0
1885594	$2\cdot3^2\cdot5\cdot7\cdot41\cdot73$	$2\cdot11\cdot13\cdot19\cdot347$	21	19	21	+2
1899206	$2\cdot19\cdot23\cdot41\cdot53$	$2\cdot3\cdot5\cdot29\cdot37\cdot59$	21	19	21	+2
1952786	$2\cdot13\cdot19\cdot59\cdot67$	$2\cdot3\cdot5\cdot7\cdot17\cdot547$	21	19	20	+1
1984814	$2\cdot5\cdot41\cdot47\cdot103$	$2\cdot3\cdot11\cdot17\cdot29\cdot61$	21	19	19	+0

TABLE 7.7. Prime-rich even values of k in $2.5\times10^5 < k < 2\times10^6$ chosen to maximise $W(k) := \omega_{\mathrm{odd}}(k) + 2\,\omega_{\mathrm{odd}}(k^2-16)$; 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

[ABLR26] Alberto Alfarano, Pablo Bianucci, Matilde N. Lalín, and Berend Ringeling. Computational data for the arithmetic of Boyd's Mahler measure conjectures, 2026. Research dataset. `https://doi.org/10.5683/SP4/LHUTSD`.

[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, Montreal, QC H4B 1R6, Canada
Email address: `pablo.bianucci@concordia.ca`

Matilde N. Lalín: Département de mathématiques et de statistique, Université de Montréal, Montreal, QC H3C 3J7, Canada
Email address: `matilde.lalin@umontreal.ca`

Berend Ringeling: Département de mathématiques et de statistique, Université de Montréal, Montreal, QC H3C 3J7, Canada
Email address: `bjringeling@gmail.com`