Journal of Number Theory 284 (2026) 104–130

Contents lists available at ScienceDirect

Journal of Number Theory

journal homepage: www.elsevier.com/locate/jnt

Computational Section

Partition-theoretic model of prime distribution

Aidan Botkin [a], Madeline L. Dawsey [b], David J. Hemmer [a], Matthew R. Just [c], Robert Schneider [a,*]

[a] Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, USA
[b] Department of Mathematics, University of Texas at Tyler, Tyler, TX 75799, USA
[c] Department of Mathematics, University of Georgia, Athens, GA 30602, USA

ARTICLE INFO

Article history:
Received 20 January 2025
Received in revised form 1 October 2025
Accepted 2 December 2025
Available online 21 January 2026
Communicated by S.J. Miller

Keywords:
Prime number distribution
Prime gaps
Integer partitions
Computational model

ABSTRACT

We make an application of ideas from partition theory to a problem in multiplicative number theory. We propose a deterministic model of prime number distribution, from first principles related to properties of integer partitions, that naturally predicts the prime number theorem as well as the twin prime conjecture. The model posits that, for $n \geq 2$,

$$p_n \;=\; 1 \;+\; 2\sum_{j=1}^{n-1}\left\lceil \frac{d(j)}{2} \right\rceil \;+\; \varepsilon(n),$$

where p_k is the kth prime number, $d(k)$ is the divisor function, and $\varepsilon(k)$ is an explicit error term that is negligible asymptotically; both the main term and error term represent enumerative functions in our conceptual model. We refine the error term to give numerical estimates of $\pi(n)$ similar to those provided by the logarithmic integral, and much more accurate than $\mathrm{li}(n)$ up to $n = 10{,}000$ where the estimates are *almost exact*. We then perform computational tests of

* Corresponding author.
E-mail addresses: abotkin@mtu.edu (A. Botkin), mdawsey@uttyler.edu (M.L. Dawsey), djhemmer@mtu.edu (D.J. Hemmer), justmatt@uga.edu (M.R. Just), robertsc@mtu.edu (R. Schneider).

https://doi.org/10.1016/j.jnt.2025.12.008

unusual predictions of the model, finding limited evidence of predictable variations in prime gaps.

1. Introduction

1.1. Model of primes from partition first principles

The prime numbers mimic a random sequence in key ways, and many successful models of prime distribution are based upon probabilistic hypotheses [43]. In his lecture notes [40], Terence Tao speaks to this point:

> "We do have a number of extremely convincing and well supported [random] models for the primes (and related objects) that let us predict what the answer to many prime number theory questions (both multiplicative and non-multiplicative) should be.... Indeed, the models are so effective at this task that analytic number theory is in the curious position of being able to confidently predict the answer to a large proportion of the open problems in the subject, whilst not possessing a clear way forward to rigorously confirm these answers!"

In this paper, a sequel to [11], we formulate a *deterministic* model of prime number distribution based on information about the integers gleaned from facts about integer partitions.[1] In short, the comparison of two multiplicative statistics defined on partitions, the norm [37] and supernorm [11], imposes constraints on the number line unrelated to integer factorizations, which we leverage to give an estimate of the nth prime gap. Computations show the model to be reasonably accurate numerically. We justify the model from partition-theoretic first principles, and we refine the error to derive an explicit prediction for the value of the nth prime number which is surprisingly close at small numbers, yielding an estimate for $\pi(n)$ (the number of primes less than or equal to n) that is *almost exact* up to $n = 10{,}000$; which predicts the prime number theorem asymptotically as n increases; and which predicts the twin prime conjecture for reasons that would be difficult to justify without ideas about partitions. The model also makes unusual predictions about local behaviors of prime gaps; we find limited computational evidence for one of these predictions. Finally, we note deficiencies in the model and suggest refinements.

Below is the central claim of our model, which we formulate from first principles about partitions. For $x \in \mathbb{R}$, let $\lceil x \rceil$ denote the usual *ceiling function*. For $x \geq 0$, let $\lfloor x \rfloor$ denote

[1] Early versions of this model were previously presented by some of the authors in [19] and the appendix of [10].

Table 1
Comparing estimates for $\pi(n)$.

n	$\pi(n)$	$n/\log n$	$\mathrm{li}(n)$	Model 1	Model 2	Model 2*
10	4	4.34...	6.16...	4	4	4
100	25	21.71...	30.12...	27	26	27
1000	168	144.76...	177.60...	184	168	171
10,000	1229	1085.73...	1246.13...	1352	1212	1233
100,000	9592	8685.88...	9629.80...	10,602	9435	9618
1,000,000	78,498	72,382.41...	78,627.54...	86,739	77,322	78,740

the *floor function*; throughout this paper *we modify the usual floor function definition to let* $\lfloor x \rfloor := 0$ *if* $x < 0$.

Partition model of prime numbers. The nth prime number $p_n,\ n \geq 1$, is modeled by setting $p_1 = 2$ and for $n \geq 2$ by using the formula

$$p_n \;=\; 1 \;+\; 2\sum_{j=1}^{n-1}\left\lceil \frac{d(j)}{2} \right\rceil \;+\; \varepsilon(n),$$

where $d(k)$ is the divisor function and $\varepsilon(k)$ is an explicit error term that is negligible by comparison. We refer to the case $\varepsilon(n) := 0$ for all $n \geq 2$ as Model 1. More nuanced consideration of the error in Model 1 leads to a refined model that we call Model 2, which for $n \geq 2$ is the case

$$\varepsilon(n) := \lfloor \pi_2(p_{n-1}) - 2\gamma(n-1) \rfloor ,$$

where $\pi_2(k)$ is the number of semiprimes less than or equal to $k \geq 1$, and $\gamma = 0.5772\ldots$ is the Euler–Mascheroni constant. To simplify calculations, we set $p_1 = 2,\ p_2 = 3$, and for $n \geq 3$ use an asymptotic approximation to the above error term,

$$\varepsilon(n) := \lfloor (n-1)\cdot(\log\log(n-1) - 2\gamma) \rfloor ,$$

where $\log x$ denotes the natural logarithm, to yield a computational model that we call Model 2*.

The reader is referred to Table 1 that gives a comparison of our models' predictions for $\pi(n)$, alongside corresponding predictions from the prime number theorem.[2]

[2] The authors are grateful to Eli DeWitt (Michigan Technological Univ.) and Alexander Walker (Univ. of Georgia) for producing Python code and computational data that advanced this project, as well as Maxwell Schneider (Univ. of Georgia) for consultation about programming and combinatorial algorithms, as undergraduate research students.

1.2. Concepts and notations

Let $\mathbb{Z}^+$ denote the *natural numbers*. Let $\mathbb{P}$ denote the *prime numbers*. Let $p_i \in \mathbb{P}$ denote the ith prime number, viz. $p_1 = 2, p_2 = 3, p_3 = 5$, etc.; we define $p_0 := 1$, and we refer to the subscript $i \in \mathbb{Z}^+$ of $p_i \in \mathbb{P}$ as the *index* of the prime number. For $n \in \mathbb{Z}^+$, we write its *prime factorization* as $n = p_1^{a_1} p_2^{a_2} p_3^{a_3} \cdots p_i^{a_i} \cdots$, $a_i \in \mathbb{Z}_{\geq 0}$, where only finitely many primes p_i have nonzero *multiplicity* (number of occurrences); we omit the factor p_i from the notation if $a_i = 0$. For $n \in \mathbb{Z}^+$, let $\pi(n)$ denote the *number of primes less than or equal to* n.

Let $\mathcal{P}$ denote the set of *integer partitions*, unordered finite multisets of natural numbers including the empty partition $\emptyset \in \mathcal{P}$ (see [3]). For a nonempty partition $\lambda \in \mathcal{P}$, we notate $\lambda = (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_r)$, with *parts* $\lambda_i \in \mathbb{Z}^+$ written in weakly decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 1$. For $\lambda \in \mathcal{P}$, let $|\lambda| \geq 0$ denote the *size* (sum of parts), let $\ell(\lambda) = r$ be the *length* (number of parts), and let $m_i = m_i(\lambda) \geq 0$ be the *multiplicity* of $i \in \mathbb{Z}^+$ as a part of partition λ. Similarly to prime factorizations, we also write partitions in *part-multiplicity* notation $\lambda = \langle 1^{m_1} 2^{m_2} 3^{m_3} \ldots i^{m_i} \ldots \rangle$, omitting any part $i \geq 1$ if $m_i = 0$. Let $p(n)$ be the *number of partitions of size* $n \geq 0$.

There are further important *partition statistics*, i.e., functions $f : \mathcal{P} \to \mathbb{Z}$ such as the partition rank, crank and others, that encode combinatorial properties in their values and often enjoy modular congruence relations and other nice behaviors [3]. Our model of prime distribution arises from consideration of two newly-defined, *multiplicative* partition statistics.

1.3. Multiplicative partition statistics and a multiplicative theory of (additive) partitions

In [32,33,35] and a series of subsequent publications, the fifth author (Schneider) introduces a *multiplicative* theory of integer partitions that parallels multiplicative number theory in many respects[3] and studies a variety of new, multiplicative partition statistics – many of them representing partition-theoretic analogues of well-known arithmetic functions like the Möbius function $\mu(n)$, the Euler phi function $\varphi(n)$, and other functions from multiplicative number theory [4]. In these works, the fifth author presents (and proves cases of) the following philosophy:

Theorems in multiplicative number theory are special cases of more general theorems in partition theory. Theorems in $\mathbb{Z}^+$ have counterpart theorems in $\mathcal{P}$, and vice versa. Facts about partitions map to facts about integers, which may be difficult to deduce without appealing to partition theory.

First applications of this philosophy were chiefly connected to the theory of Dirichlet series through the study of new classes of partition zeta functions [28,32,36], and to q-

[3] Inspired by Alladi-Erdős [1], Andrews [2], Granville [16], Granville-Soundararajan [17], Ono [27] and Zagier [44].

series and quasimodular forms through work on the q-bracket of Bloch and Okounkov [33,34]. In [29,30], Ono, Wagner and the fifth author make a computational application using partition-theoretic q-series to give limit formulas for arithmetic densities that are comparable to those using Dirichlet series; this is generalized in [38]. A particularly stunning application of this philosophy is made by Craig, Ono and van Ittersum [8], in which the partition-theoretic coefficients of certain quasimodular forms are proved to be intrinsically connected to the prime numbers.[4]

In [33,35], a *partition multiplication* operation is introduced; this comes with a concept of "subpartitions" analogous to integer divisors and a theory of partition Dirichlet convolution.

Definition 1.1. For two partitions $\lambda, \gamma \in \mathcal{P}$, we define the *partition product* $\lambda \cdot \gamma \in \mathcal{P}$ (or simply $\lambda\gamma$) to be the partition obtained by concatenating the parts of λ and γ (and then reordering by size to align with notational convention). For instance, $(5,3,2)\cdot(4,3,1,1) = (5,4,3,3,2,1,1)$.[5]

A new multiplicative partition statistic is introduced in [32] that is central to the study of partition zeta functions, a multiplicative analogue of the size $|\lambda|$ called the *norm* $N(\lambda)$ of $\lambda \in \mathcal{P}$.

Definition 1.2. The *norm* of an integer partition $\lambda = (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_r)$ is the product of its parts:

$$N(\lambda) \ := \ \lambda_1\lambda_2\lambda_3\cdots\lambda_r \ = \ 1^{m_1}2^{m_2}3^{m_3}\cdots i^{m_i}\cdots \in \mathbb{Z}^+; \tag{1.1}$$

we define $N(\emptyset) := 1$ (it is an empty product).

We note that $N(\lambda\gamma) = N(\lambda)N(\gamma)$ for $\lambda, \gamma \in \mathcal{P}$; see [37] for more about the partition norm. Along these lines, the multiplicative and additive branches of number theory enjoy further analogies.[6]

1.4. Isomorphism between partitions and natural numbers

In [11], the present authors Dawsey, Just, and Schneider define another multiplicative partition statistic, the *supernorm* $\widehat{N}(\lambda)$.

Definition 1.3. The *supernorm* of an integer partition $\lambda = (\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_r)$ is the product

$$\widehat{N}(\lambda) \ := \ p_{\lambda_1}p_{\lambda_2}\cdots p_{\lambda_r} \ = \ 2^{m_1}3^{m_2}5^{m_3}\cdots p_i^{m_i}\cdots \in \mathbb{Z}^+, \tag{1.2}$$

[4] Follow-up works such as [7,15] apply similar ideas to prime powers.

[5] Equivalently, one sums the multiplicities of corresponding parts in the two partitions.

[6] For further references at the intersection of partitions and multiplicative number theory, see e.g. [1,20,25].

where $p_i \in \mathbb{P}$ is the ith prime number, $i \geq 1$, and $m_j = m_j(\lambda) \geq 0$ as above; we define $\widehat{N}(\emptyset) := 1$.

We note that $\widehat{N}(\lambda\gamma) = \widehat{N}(\lambda)\widehat{N}(\gamma)$ for $\lambda, \gamma \in \mathcal{P}$. In [11], it is proved that $\widehat{N}$ induces an isomorphism of monoids.

Theorem *(Dawsey–Just–Schneider). The supernorm map $\widehat{N} \colon \mathcal{P} \to \mathbb{Z}^+$ is an isomorphism between monoid $(\mathcal{P}, \ \cdot\ )$ with partition multiplication and monoid $(\mathbb{Z}^+, \ \cdot\ )$ with integer multiplication.*

It is the central theme of [11] that *the supernorm translates facts about partitions to analogous facts about integers, with partition parts mapping to prime factors, and vice versa.* For example, in [11], bijections between integers with certain prime factorizations, and partition formulas for natural densities of subsets of prime numbers, are proved. In a follow-up note [39], the fifth author extends partition multiplication to the set $\mathcal{O}$ of overpartitions (see e.g. [6]) and proves that an extension $\widehat{N}_{\mathcal{O}} \colon \mathcal{O} \to \mathbb{Q}^+$ of the supernorm to overpartitions induces a *group* isomorphism between $(\mathcal{O}, \ \cdot\ )$ with partition multiplication and $(\mathbb{Q}^+, \ \cdot\ )$ with rational multiplication. In [21], J. Lagarias proves further, striking properties of the supernorm that uniquely characterize the map $\widehat{N}$ in a lattice-theoretic context. Moreover, Lagarias and Sun prove interesting asymptotic and statistical relations using sums of the reciprocal of the supernorm over different finite subsets of $\mathcal{P}$ in [22].

1.5. Comparing the norm and supernorm

Immediately one wonders about the magnitude of the supernorm compared to the partition norm; some analysis to this effect is undertaken in [11]. Here are some easy relations. First, a partition λ with no part equal to 1 respects the inequalities

$$\ell(\lambda) \ \leq\ |\lambda| \ \leq\ N(\lambda) \ \leq \widehat{N}(\lambda). \tag{1.3}$$

A simple observation about the supernorm versus the norm that one can make from direct computation is as follows: *partitions with greater lengths or having many parts equal to* 1 *have supernorms that are much larger than their norms; furthermore, partitions with fewer parts and no part equal to* 1 *have supernorms closer to their norms.* This empirical observation – that seems evident but we have not proved – is compatible with Rosser's theorem [31], which says for $n \geq 1$,

$$p_n \ >\ n \log n; \tag{1.4}$$

thus, for all $\lambda \in \mathcal{P}$,

$$\widehat{N}(\lambda) \ >\ 2^{m_1(\lambda)} \prod_{i\geq 2} (i \log i)^{m_i(\lambda)} \ =\ N(\lambda) \cdot 2^{m_1(\lambda)} \prod_{i\geq 2} (\log i)^{m_i(\lambda)}. \tag{1.5}$$

The right-hand side of (1.5) is closer to $N(\lambda)$ when λ has fewer parts and no part equal to 1.

Moreover, in [11] it is proved that if λ has no part equal to 1, then

$$N(\lambda) \leq \widehat{N}(\lambda) \leq N(\lambda)^{\log 3/\log 2}, \tag{1.6}$$

noting $\log 3/\log 2 = 1.5849\ldots$, so the supernorm of the partition is not far off in order of magnitude from the norm. The inequalities in (1.6) tell us *partitions having fixed norm and no part equal to* 1 *are mapped into a relatively small interval on the number line by the supernorm.*

We observe further that if λ is an unrestricted partition of size $|\lambda| = n$ and p_n is the nth prime number, then the supernorm falls in the interval

$$p_n \leq \widehat{N}(\lambda) \leq 2^n, \tag{1.7}$$

with equality on the left-hand side when $\lambda = \langle n^1 \rangle$, and equality on the right when $\lambda = \langle 1^n \rangle$; this is proved in [11]. By similar reasoning, if λ is a partition with no part equal to 1, $|\lambda| = n$, then

$$p_n \leq \widehat{N}(\lambda) \leq 3^{n/2}, \tag{1.8}$$

with equality on the right when n is even and $\lambda = \langle 2^{n/2} \rangle$. The above inequalities suggest *partitions of fixed size are mapped across a relatively large interval on the number line by the supernorm*, by comparison with partitions of fixed norm.

Combining the inequalities above, we see for λ a partition with no part equal to 1,

$$p_{|\lambda|} \;\leq\; \widehat{N}(\lambda) \;\leq\; N(\lambda)^{\log 3/\log 2}. \tag{1.9}$$

One further inequality one can observe computationally, which is proved by J.E. Cohen [5] as a consequence of Rosser's theorem,[7] is that for $\lambda \in \mathcal{P}$ such that $N(\lambda) \geq 5$,

$$p_{N(\lambda)} \leq \widehat{N}(\lambda). \tag{1.10}$$

Inequality (1.10) says that for $n \geq 5$, *partitions of fixed norm equal to n have supernorms lying above p_n* (similarly to fixed size above).

From these inequalities, the central observation of this paper comes into view: *comparison of the partition norm and supernorm imposes constraints on the prime numbers.* For instance, from (1.9) one can deduce an upper bound for the nth prime number.

Proposition 1.4. *For $n \geq 2$, we have that*

$$p_n \;\leq\; n^{\log 3/\log 2}.$$

[7] The authors are grateful to Abhimanyu Kumar for pointing us to Cohen's proof of this inequality.

Proof. Take $\lambda = (n),\ n \geq 2$; then substitute $|\lambda| = N(\lambda) = n$ on the left and right sides of (1.9). □

In Section 2, we formulate a model of primes based on the above observations about partitions.

2. Partition model of prime distribution

2.1. Partition-theoretic model of prime gaps

We use our preceding comparisons of the partition norm and supernorm to formulate a model of prime numbers.

While the set of primes is highly enigmatic in number theory and known for its seemingly random distribution, from the perspective of the supernorm, prime numbers have somewhat regular behavior: the nth prime p_n represents the supernorm $\widehat{N}(\lambda)$ of the partition $\lambda = (n)$ with a *single part* $n \in \mathbb{Z}^+$ (recalling $\emptyset$ maps to 1). The sequence of primes is the image of the sequence of natural numbers under the map $\widehat{N}$, while composite integers fill in the gaps in a more complicated way.

From this perspective, the complexities of prime distribution are a manifestation of the complex proliferation of *partitions with multiple parts*, which are mapped by the supernorm to *gaps between primes*. Thus prime gaps have a combinatorial interpretation: they are the images of certain subsets of $\mathcal{P}$ under the supernorm map. This implies the following statement.

Proposition 2.1. *Measuring prime gaps is equivalent to enumerating partitions that map into the respective prime gaps under the supernorm $\widehat{N}$:*

$$p_{n+1} - p_n \quad = \quad \#\left\{\lambda \in \mathcal{P} \ : \ p_n \leq \widehat{N}(\lambda) < p_{n+1}\right\}.$$

Proof. This is immediate since the supernorm is a bijection between $\mathcal{P}$ and $\mathbb{Z}^+$. □

Our goal based on Proposition 2.1 is to identify partitions that map into each prime gap so we can count them. This goal suggests a potentially workable heuristic.

Heuristic. Make an educated guess as to which partitions map into the nth prime gap under the supernorm $\widehat{N}$, in order to estimate the value of $p_{n+1} - p_n$.

Below, we will refine this prime gap heuristic from elementary considerations to formulate an initial, if overly simplistic, partition model of prime numbers that agrees with the facts quite well. We then address sources of error to improve the model significantly.

2.2. Simplifying assumption

We wish to simplify the problem as much as possible. The goal is to estimate the length of the interval $[p_n, p_{n+1})$ with high accuracy for each $n \geq 2$. That is, we want to enumerate partitions we expect to map to an interval just above p_n under the supernorm. We summarize observations we made in Section 1.5 that may aid us in identifying such partitions:

(1) Partitions λ with a smaller number $\ell(\lambda)$ of parts and with no part equal to 1 should have supernorms closer to their norms in magnitude, based on the considerations preceding (1.4).
(2) Partitions λ with fixed norm $N(\lambda) = n$ and with no part equal to 1 should have supernorms of roughly comparable magnitudes to each other, based on consideration of (1.6).
(3) Partitions λ with fixed norm $N(\lambda) = n \geq 5$ and with no part equal to 1 respect the inequality $p_n \leq \widehat{N}(\lambda) \leq \ n^{\log 3/\log 2}$, based on consideration of (1.9) and (1.10).

Taken together, the preceding observations produce a qualitative model in our minds: *The number line just above p_n is dominated by the images under $\widehat{N}$ of partitions of norm equal to n, having no part equal to* 1*, and possessing a small number of parts.* We note partitions with no 1's yield *odd* supernorm values; for each odd value, we must also count the even number that follows. Furthermore, the norm-n partitions with *many* parts are mapped by $\widehat{N}$ closer to the upper bound $n^{\log 3/\log 2}$ and do not map into the gap $[p_n, p_{n+1})$, but rather give rise to a lengthy "tail" of leftover integers that contribute to subsequent prime gaps more diffusely, perhaps somewhat randomly. The interval $[p_n, p_{n+1})$ also includes integers arising from the "tails" of subsets of partitions with norms less than n; we anticipate their contribution is negligible asymptotically.

To translate this qualitative vision into a computational model one can test and use to make predictions, we shall assume the following extreme simplification to address the observations above.[8]

Simplifying Assumption 1. Assume for simplicity that all odd integers in the interval $[p_n, p_{n+1})$, $n \geq 2$, are the images of partitions with norm equal to n, with no part equal to 1, having one or two parts, under the supernorm $\widehat{N}$. Then

$$p_{n+1} - p_n \quad = \quad 2 \cdot \# \{\lambda \in \mathcal{P} \ : \ N(\lambda) = n,\ m_1(\lambda) = 0,\ \ell(\lambda) = 1 \text{ or } 2\}. \tag{2.1}$$

Remark 2.1. That is, we will assume the images of the above partitions dominate the number line immediately above p_n.

[8] In [10, Appendix], the second, fourth, and fifth authors give more detailed discussion of this assumption.

Integers having two prime factors are called *semiprimes*. We expect this simplification yields an *underestimate* of prime gaps from the incorrect assumption all odd numbers are prime or semiprime.

Now observe that the partitions of norm $n \geq 2$, with no part equal to 1 and having one or two parts, are precisely the set consisting of the partition (n) into one part, together with partitions (d_1, d_2) where $d_1 d_2 = n,\ d_1 \geq d_2$; that is, partitions whose parts are pairs of divisors of n. There are $\lceil d(n)/2 \rceil$ such divisor pairs where $d(k)$ is the *divisor function*, setting $d(0) := 0$,[9] including $d_1 = n, d_2 = 1$, which we associate to the partition (n). Then Simplifying Assumption 1 translates to the relation

$$p_{n+1} - p_n \;=\; 2\left\lceil \frac{d(n)}{2} \right\rceil, \tag{2.2}$$

where $\lceil x \rceil$ is the ceiling function.

We refer the reader to Table 2 to see (2.2) does correctly predict most of the prime gaps between 1 and 100. Computations show this near-correctness diminishes as n increases, but as we will prove, equation (2.2) is compatible with the prime number theorem asymptotically.

2.3. Initial model

The following formulation is suggested by discrepancies between the partition norm and supernorm noted in Section 1.5, which led to Simplifying Assumption 1 above. Since $p_n = 1 + \sum_{k=0}^{n-1}(p_{k+1} - p_k)$ by telescoping series, defining $p_0 := 1$ as we did above, then replacing each summand $p_{k+1} - p_k$ by the estimate $2\lceil d(k)/2 \rceil$ from (2.2) leads us to a model of prime numbers.

Model 1. The prime numbers $p_1, p_2, p_3, \ldots,$ can be modeled by the sequence having initial value $p_1 = 2$ and for $n \geq 2$ having the values

$$p_n \;=\; 1 \;+\; 2\sum_{k=1}^{n-1} \left\lceil \frac{d(k)}{2} \right\rceil.$$

Remark 2.2. For computational ease, note $2\lceil d(k)/2 \rceil = d(k)$ for all $k \neq m^2, m \in \mathbb{Z}^+$. On the other hand, if k is a perfect square then $2\lceil d(k)/2 \rceil = d(k) + 1$. Thus for $N \geq 1$, we have that

$$2\sum_{k=1}^{N} \left\lceil \frac{d(k)}{2} \right\rceil \;=\; \sum_{k=1}^{N} d(k) \;+\; \left\lfloor \sqrt{N} \right\rfloor \;=\; \sum_{k=1}^{N} \left\lfloor \frac{N}{k} \right\rfloor \;+\; \left\lfloor \sqrt{N} \right\rfloor, \tag{2.3}$$

[9] We define $d(0) := 0$ for our purposes here, as there are no positive integers d_1, d_2, such that $d_1 d_2 = 0$.

Table 2
Comparing actual prime gaps to predictions from Model 1; we highlight entries where the prediction for the nth prime gap is off.

n	p_n	$p_{n+1} - p_n$ (Actual)	$p_{n+1} - p_n$ (Model 1)
1	2	1	1
2	3	2	2
3	5	2	2
4	7	4	4
5	11	2	2
6	13	4	4
7	17	2	2
8	19	4	4
9	23	6	4
10	29	2	4
11	31	6	2
12	37	4	6
13	41	2	2
14	43	4	4
15	47	6	4
16	53	6	6
17	59	2	2
18	61	6	6
19	67	4	2
20	71	2	6
21	73	6	4
22	79	4	4
23	83	6	2
24	89	8	8
25	97	4	4

by well-known methods related to the summatory function of $d(k)$ [4]. This yields an equivalent formula in Model 1 that does not require explicitly computing the divisor function:

$$p_n \;=\; 1 + \sum_{k=1}^{n-1} \left\lfloor \frac{n-1}{k} \right\rfloor \;+\; \left\lfloor \sqrt{n-1} \right\rfloor . \tag{2.4}$$

The first ten terms of this sequence of estimated values for p_n are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 27,$$

and from here Model 1 does not give the correct sequence of prime values. However, the model does *almost* give the correct sequence of *prime gaps* at small numbers; see Table 2.

One can estimate $\pi(n)$ by counting the primes up to n predicted by the model. Table 1 and Table 2 display that Model 1 does capture aspects of prime distribution qualitatively: at small numbers, it emulates the unpredictable prime gaps reasonably well, and at large numbers, it yields numerical estimates for $\pi(n)$ that are comparable to the prime number theorem. Note from Table 1 that Model 1 evidently leads to an *overestimate* for $\pi(n)$.[10]

[10] We interpret this from our premises as a result of the model's representing an underestimate of prime gaps.

In the next subsection, we discuss explicit predictions of Model 1 and try to interpret them.

Remark 2.3. We note that J. N. Gandhi proves a formula for the nth prime number in [13], which represents a base-2 version of the sieve of Eratosthenes [14].[11]

2.4. Predictions from Model 1

Below we show that Model 1 implies the main asymptotic term of the prime number theorem and produces unusual predictions about local fluctuations in prime gaps, including the infinitude of twin primes as a special case.

The *prime number theorem* was conjectured by both Gauss and Legendre in the 1790s based on newly published tables of primes, and was known as the “prime number conjecture” for a century, until Hadamard and de la Vallée Poussin proved it independently at the very end of the nineteenth century [43]. The prime number theorem is usually expressed as an asymptotic estimate for $\pi(n)$,

$$\pi(n) \;\sim\; \frac{n}{\log n} \tag{2.5}$$

for $n \geq 2$ as n increases, where $\log x$ denotes the natural logarithm; or by the asymptotically equivalent (and evidently more accurate) estimate provided by the *logarithmic integral* $\mathrm{li}(n)$:

$$\pi(n) \;\sim\; \mathrm{li}(n) := \int_2^n \frac{dt}{\log t}. \tag{2.6}$$

The prime number theorem is equivalently formulated as an asymptotic estimate for the nth prime:

$$p_n \;\sim\; n \log n. \tag{2.7}$$

The estimate (2.7) follows from (2.5) by observing that $\pi(p_n) = n \sim p_n/\log p_n \sim p_n/\log n$. This final formulation of the prime number theorem is suggested naturally by Model 1. The first prediction of the model is a statement of fact.

Prediction 1 *(Theorem). Model 1 predicts the prime number theorem's estimate for the nth prime,*

$$p_n \;\sim\; n \log n \quad \textit{as} \quad n \to \infty.$$

[11] The authors are grateful to Ken Ono for pointing us to Gandhi's formula.

Proof. It follows from Simplifying Assumption 1 that

$$p_n = 1 + \sum_{0 \le i \le n-1} (p_{i+1} - p_i) \sim \sum_{1 \le i \le n-1} d(i) = (n-1)\log(n-1) + (n-1)(2\gamma - 1) + O(\sqrt{n}) \tag{2.8}$$

where $\gamma = 0.5772\ldots$ is the Euler-Mascheroni constant; the final equality is due to Dirichlet [4]. The far right-hand side of (2.8) is asymptotic to $n \log n$ as n increases. □

Remark 2.4. We note there is a second-order error term of magnitude $n \log \log n$ implied by the prime number theorem, arising from $n = \pi(p_n) \sim p_n / \log p_n$, since $\log p_n \sim \log(n \log p_n) \sim \log n + \log \log n$, that is missing from Prediction 1. We address this missing error in Section 4.

An unusual feature of Model 1 is that it suggests *predictable local behaviors of prime gaps.*

Prediction 2. *For $n \ge 2$, Model 1 predicts the nth prime gap is larger or smaller, depending on if n has a larger or smaller number of divisors, respectively.*

Proof. This is immediate from (2.2), as $d(n)$ effectively controls the prime gap in this model. □

Prediction 2 is counterintuitive: it says that *the indices of primes influence the prime gaps.* Table 2 gives evidence of this up to $n = 100$, where the majority of prime gaps match the predicted values exactly, but further computations indicate this exact matching does not hold as n increases. On the other hand, it follows from the prime number theorem that the average number of prime gaps continues to match Model 1's predictions asymptotically. Interpreted strictly, Model 1 says the nth prime gap should be close to $d(n)$; Prediction 2 is a more probabilistic interpretation.

The twin prime conjecture is a natural consequence of Model 1.

Prediction 3. *Model 1 predicts twin prime pairs occur at prime-indexed primes, i.e., at primes p_n such that $n \in \mathbb{P}$. Moreover, Model 1 predicts the set of twin primes is infinite.*

Proof. This prediction focuses on a special case of Model 1, the case in which $d(n) = 2$ if and only if $n \in \mathbb{P}$. Since prime-indexed primes are associated to twin prime pairs in Model 1, the predicted infinitude of twin primes follows from the infinitude of the primes proved by Euclid. □

Known observations on finite intervals are suggestive of the infinitude prediction, although as of this writing, the twin prime conjecture is undecided. However, since we expect that Model 1 is an underestimate of most prime gaps, it cannot claim to model twin primes with high accuracy; the model does not claim the precision to definitively predict a gap of minimal size 2, ruling out possible contributions from partitions with more than two parts mapping into the gap under $\widehat{N}$. We wonder if Prediction 3 can be interpreted probabilistically like Prediction 2: *prime-indexed primes are more likely to belong to twin prime pairs by comparison with arbitrary prime numbers.*

In Section 3, we make a preliminary check of these predictions.

Remark 2.5. One can formulate further, similar predictions from Model 1. For example, consider a prime number p_m whose index $m \in \mathbb{P}$ is itself the first member of a pair of twin primes, i.e., m and $m+2$ are both prime. The model predicts p_m will be the first member of a prime 4-tuplet, meaning that $p_m,\ p_{m+1}$ are twin primes and $p_{m+2},\ p_{m+3}$ are also twin primes (see e.g. [12]).

2.5. Critique of Model 1

As noted previously, while Model 1 emulates certain features of prime distribution, we expect *a priori* that it represents an underestimate of prime gaps in general and an overestimate for $\pi(n)$, since the model does not account for odd numbers with more than two prime factors; this expectation is supported by our data. Moreover, the model is missing a second-order error term of size $n \log\log n$ by comparison with the prime number theorem's estimate for p_n.

While the model impressed itself, so to speak, on the authors from observations about the partition norm and supernorm, the considerations preceding (1.4) involve an empirical observation that is not proved rigorously, although we noted it is compatible with Rosser's theorem. Furthermore, we have not proved (or observed) that *even one* partition with norm n should necessarily map into the nth prime gap under the supernorm; this is an assumption justified by its predictive success. Model 1 also does not suggest that composite integers should respect anything like the correct integer ordering; nor does it account for constraints on primes imposed by multiplication.

We test Predictions 2 and 3 in Section 3. In Section 4, we examine the error in Model 1 and make further simplifying assumptions that lead to a computational model with greater accuracy.

3. Testing the model

3.1. Testing predictions from Model 1

Tables 1 and 2 and Prediction 1 provide reasonable support for Model 1. Predictions 2 and 3 place unusual emphasis on the factorizations of the *indices* of prime numbers. We give these unusual predictions a preliminary check.

3.2. Methods

We utilize the high-performance computing shared facility at Michigan Technological University to check Predictions 2 and 3. For calculations we use Wolfram Mathematica computer algebra system and Python programming language. The authors made use of OpenAI GPT-4o large language model software to assist with writing Python code to generate experimental data exclusively in Sections 3.3 and 3.4 below; we checked and revised this code as needed and take full responsibility for the results. We will share our computer code and data upon request.

3.3. Computational test of Prediction 2

Prediction 2 says the nth prime gap is larger or smaller depending on the size of $d(n)$. As we noted above, Table 2 is consistent with this prediction because $p_{n+1} - p_n$ is equal to $2\lceil d(n)/2\rceil$ *exactly* for most of the prime gaps in the table. However, further computations reveal the exact equalities decrease in frequency as n increases.

We use a less exact method to check this prediction. It follows from the prime number theorem that the *average order* of the nth prime gap is asymptotic to $\log p_n \sim \log n + \log\log n$ as $n \to \infty$. As is standard (see e.g. [26]), we define the *merit* $M(n)$ of the nth prime gap to be the ratio

$$M(n) \;:=\; \frac{p_{n+1} - p_n}{\log p_n}. \tag{3.1}$$

When the merit is greater or less than one, the prime gap is larger or smaller than average, respectively. Now, the average order of $d(n)$ is asymptotic to $\log n$, and likewise for $2\lceil d(n)/2\rceil$ that is equal to $d(n)$ except at perfect squares when it equals $d(n)+1$. For $n \geq 2$, let us define the *Model 1 merit* $M_1(n)$ to be the analogous ratio with respect to the nth *modeled* prime gap,

$$M_1(n) \;:=\; \frac{d(n)}{\log n}; \tag{3.2}$$

we omit the ceiling function for computational ease since $d(n) = 2\lceil d(n)/2\rceil$ almost always. When the Model 1 merit is greater or less than one, the divisor function is larger or smaller than average.

To give a preliminary check of Prediction 2 – whether prime gaps tend to increase or decrease with $d(n)$ – we count how often $M(n)$ and $M_1(n)$ are simultaneously greater or less than one.

Result of computation. We compute $M(n)$ and $M_1(n)$ for $n \leq 1{,}000{,}000$. We find $M(n) > 1$ in 36.01% of instances and $M(n) < 1$ in 63.99% of instances. We find $M_1(n) > 1$ in 37.94% of instances and $M_1(n) < 1$ in 62.06% of instances. Finally, we find $M(n)$ and $M_1(n)$ are simultaneously greater or less than one in 53.34% of instances.

Remark 3.1. The similarity between corresponding $M(n)$ and $M_1(n)$ percentages is noteworthy. For instance, one expects $M(n) < 1$ to occur often since smaller prime gaps are more abundant than larger gaps up to n (see e.g. [41]). One also expects $M_1(n) < 1$ often since the *normal* order of $d(n)$ is well known to be asymptotic to $\log\log n = o(\log n)$ as $n \to \infty$. But we are not aware of a well-known reason $p_{n+1} - p_n$ should be smaller than average with similar frequency to the divisor function. Perhaps an analysis along the lines of that given in [41] would provide an explanation.

Up to one million, a slight majority of prime gaps increase or decrease with the divisor function.

3.4. Computational test of weak version of Prediction 3

In Section 2.4 we make a probabilistic conjecture based on Prediction 3, that a prime-indexed prime is more likely to begin a twin prime pair than is an arbitrary prime number. Table 1 supports this weak version of Prediction 3 up to $n = 100$; there are 6 out of 8 twin prime pairs correctly predicted in the table (but these exact results do not continue as n increases) and an additional three are predicted erroneously.

Furthermore, one can deduce from Cramér's random model of primes that the number of twin prime pairs less than or equal to n is asymptotic to $n/(\log n)^2$ [43]. The number of prime-indexed primes less than or equal to n is $\pi(\pi(n)) \sim \pi(n)/\log\pi(n) \sim n/(\log n)^2$ by the prime number theorem so, in fact, *we do expect twin prime pairs and prime-indexed primes to be approximately equinumerous at large numbers.* How often do twin prime pairs and prime-indexed primes co-occur?

Let us refer to a prime gap whose first member has prime index as a *prime-indexed prime gap*, and to a twin prime pair whose first member has prime index as a *prime-indexed twin prime pair*.

We run a few cursory experiments to explore statistics related to Prediction 3. First, we give a simple "greater-or-less" test like the one in Section 3.3 related to the merit $M(n)$. Model 1 predicts $M(n) = 2/\log p_n < 1$ every time n is prime since the nth prime gap is $2\lceil d(n)/2\rceil$ in the model.

Result of computation. We only compute merits of prime-indexed prime gaps up to one million, i.e., we compute $M(p)$ over primes $p \le \pi(1{,}000{,}000) = 78{,}498$. We find $M(p) > 1$ in 37.39% of instances and $M(p) < 1$ in 62.61% of instances. Finally, up to one million we find 10.59% of prime-indexed prime gaps are also twin prime pairs, and 9.98% of twin prime pairs begin with prime-indexed primes. Data suggests these final percentages decrease as $n \to \infty$; see Table 3.

This result is inconclusive. We find $M(n) < 1$ in the majority of cases, a result compatible with Prediction 3. However, our computational results for $M(n)$ restricted to prime values of n split into almost the same "greater-than" and "less-than" proportions

Table 3
Total number of twin prime pairs vs. those beginning with prime-indexed primes.

n	# Twin prime pairs $\leq n$	# Prime-indexed twin prime pairs $\leq n$
10	2	2
100	8	6
1000	35	12
10,000	205	30
100,000	1224	154
1,000,000	8169	816

as the computations over all n in Section 3.3; apparently restricting to prime indices n does not increase the $M(n) < 1$ proportion.

3.5. Second test of weak version of Prediction 3

Next, we directly compare the probability that a prime-indexed prime begins a twin prime pair with the probability that an arbitrary prime number does so. Define

$$\text{Prob}_1(n) := \frac{\#\{\text{prime-indexed primes } p \leq n \text{ such that } p+2 \text{ is also prime}\}}{\#\{\text{prime-indexed primes } p \leq n\}}, \tag{3.3}$$

the probability that a random prime-indexed prime $\leq n$ begins a twin prime pair, and define

$$\text{Prob}_2(n) := \frac{\#\{\text{primes } p \leq n \text{ such that } p+2 \text{ is also prime}\}}{\pi(n)}, \tag{3.4}$$

the probability that a random unrestricted prime $\leq n$ begins a twin prime pair. We compute

$$P(n) := \frac{\text{Prob}_1(n)}{\text{Prob}_2(n)} \tag{3.5}$$

as n increases. By probabilistic reasoning, we expect the ratio $P(n)$ to be around one: prime-indexed primes and arbitrary primes should have the same probability of being twin primes.

Result of computation. At small numbers, $P(n)$ is almost always greater than one with minor oscillations in the value locally, except for infrequent larger oscillations dipping below one. From about $n = 10{,}000{,}000$ up to at least 10 billion, $P(n)$ is strictly greater than one. See Fig. 1.

This result is compatible with Prediction 3 but we can see another possible explanation for it: *on average, primes become sparser and twin prime pairs become rarer as n grows.* The typical prime-indexed prime less than or equal to n is on average smaller than the

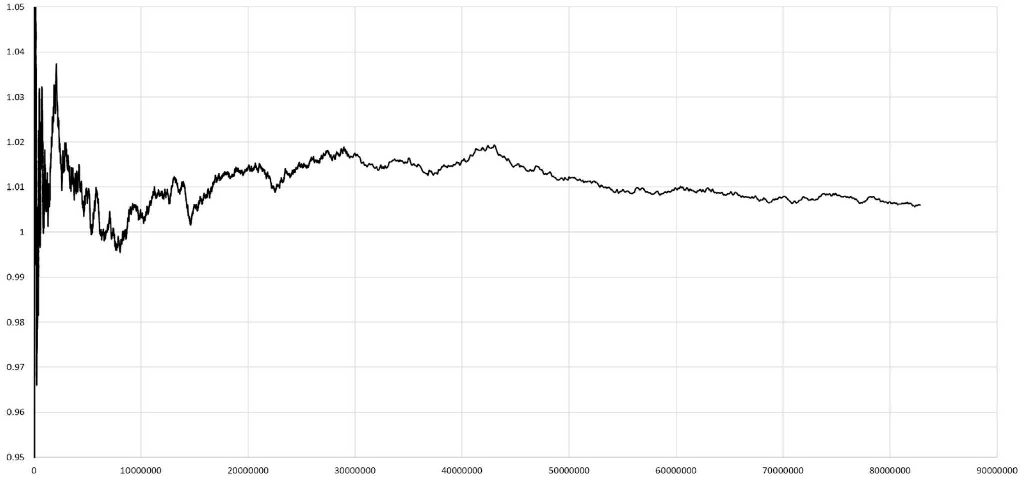

Fig. 1. Plot of $P(n)$ up to $n = 100{,}000{,}000$.

typical arbitrary prime on the same interval, and thus may be more likely to be the first of a twin prime pair not because it is prime-indexed, but simply because it is smaller.

In an attempt to remove this bias related to small numbers, we look at prime-indexed primes versus arbitrary primes on the interval $[n, 2n)$. Let $Q(n)$ denote the analogous ratio to $P(n)$ in (3.5) with both probabilities evaluated on the interval $[n, 2n)$ instead of $[1, n]$.[12]

Result of computation. We compute $Q(n)$ up to $n = 500{,}000$. We find $Q(n)$ oscillates irregularly around one. We note a tendency for $Q(n)$ to be greater than one up to $n = 80{,}000$ but not on larger intervals, where $Q(n)$ appears to approach one asymptotically as n increases.

We conclude from the computational experiments in Section 3.4 and in this subsection that Prediction 3 may hold at small numbers,[13] but does not appear to hold as n increases.

4. Improvements to the model

4.1. Refining the error term

We now turn our attention from formulating a conceptual model of prime distribution to improving our computational model and its numerical estimates for $\pi(n)$. Recall that

[12] This $Q(n)$ test was suggested to the authors by Michael Filaseta. We are grateful to Filaseta for sharing invaluable advice on computing probabilities such as these (Private communication, April 12–13, 2025).

[13] However, we are not sure if this results from the prime indices or from probabilistic anomalies at small numbers.

we define a semiprime to be an integer having exactly two prime factors. Let $\pi_k(n)$ denote the number of positive integers less than or equal to $n \geq 1$ having exactly $k \geq 0$ prime factors (so $\pi_1(n) = \pi(n)$ and $\pi_0(n) = 1$ for all $n \geq 1$). All the integers up to n are enumerated by

$$n \;=\; 1 + \pi(n) + \pi_2(n) + \pi_3(n) + \cdots + \pi_k(n) + \dots \tag{4.1}$$

with only finitely many nonzero terms, depending on n. Substituting $n \mapsto p_n$ in (4.1) gives

$$p_n \;=\; 1 + n + \pi_2(p_n) + \pi_3(p_n) + \dots . \tag{4.2}$$

This provides an explicit formula for the nth prime gap in terms of values $\pi_k(p), p \in \mathbb{P}$:

$$p_{n+1} - p_n \;=\; 1 + [\pi_2(p_{n+1}) - \pi_2(p_n)] + [\pi_3(p_{n+1}) - \pi_3(p_n)] + \dots . \tag{4.3}$$

Now, (4.2) and (4.3) are not conceptually compatible with Model 1; the equations track factorizations of even as well as odd integers in the prime gap, while Simplifying Assumption 1 only allows for enumerating odd semiprimes in the image of the supernorm and doubling the number. But these identities can inform corrections to Model 1. It is an estimate due to Landau [23] that for $n > 1, k \geq 1$, as $n \to \infty$ we have

$$\pi_k(n) \;\sim\; \frac{n\,(\log\log n)^{k-1}}{\log n\ (k-1)!}. \tag{4.4}$$

The $k = 1$ case of (4.4) is the prime number theorem. It follows for $k \geq 1$ that[14]

$$\pi_k(p_{n+1}) - \pi_k(p_n) \;\sim\; \frac{(\log\log n)^{k-1}}{(k-1)!}. \tag{4.5}$$

For fixed $n \geq 3$, this final ratio approaches zero as $k \to \infty$. Therefore, we expect integers with $k = 3$ prime factors are the next most populous set in a prime gap, apart from semiprimes. To improve our model, we should model the contributions of integers with three prime factors.

However, for conceptual and computational simplicity in our model, we want to avoid considering factorizations with many prime components. We posit a second simplifying assumption that avoids having to perform factorizations involving more than two prime factors, based on the following elementary observation: *every integer less than p_{n+1} having three prime factors is the image of a semiprime less than p_n multiplied by a prime.* We will assume for simplicity that *the converse also holds*, as an approximate approach to modeling the contribution of integers with three factors.

[14] Noting $\log p_n \sim \log(n \log n) = \log n + \log\log n \sim \log n \sim \log(n+1) \sim \log p_{n+1}$ as $n \to \infty$.

Define a *k-almost prime* to be an integer having exactly $k \geq 2$ prime factors (see e.g. [18]).

Simplifying Assumption 2. Assume for simplicity that every semiprime less than p_n maps into the interval $[2,\ p_{n+1}),\ n \geq 2$, under multiplication by some prime number, and that the 3-almost primes resulting from these products are not already enumerated by Model 1. By (4.4), as n increases this yields an estimated contribution of

$$\pi_2(p_n) \ \sim \ \pi(p_n) \log\log p_n \ \sim \ n \log\log n \tag{4.6}$$

integers having three prime factors in the interval $[2, p_{n+1})$, in addition to the odd semiprimes and their even "doubles" enumerated by Model 1.

We count these integers with three prime factors in our updated model. *Simplifying Assumption 2 fills in the missing error of order $n \log\log n$ noted previously.* Simplifying Assumption 2 likely produces an overestimate, but we have not proved this; only semiprimes less than $p_{n+1}/2 < p_n$ can map into the interval $[2, p_{n+1})$ via multiplication but some semiprimes might produce more than one 3-almost prime that is less than p_{n+1}, through multiplication by different primes.

We make one further simplifying assumption that is solely computational, and as such, represents an *ad hoc* revision – yet it leads to a significantly more accurate computational estimate for $\pi(n)$ that is *almost exact* at small numbers. One may want to replace the second summation in (2.8) with an integral. Note the asymptotic estimate on the right side of (2.8) can be expressed as

$$p_n \ \sim \ \int_0^{n-1} \log t \ dt \ + \ 2\gamma(n-1). \tag{4.7}$$

This integral estimate is simpler, and unaffected asymptotically, if we delete the $2\gamma(n-1)$ term.[15]

Simplifying Assumption 3. Assume for simplicity that subtracting $2\gamma(n-1)$ from the model will not hurt the reasonable accuracy of the model at small numbers and will have negligible impact on the estimate for p_n asymptotically.

Combining Simplifying Assumptions 2 and 3 above with the statement of Model 1, we posit a refined model. We add a correction term $\lfloor \pi_2(p_{n-1}) - 2\gamma(n-1) \rfloor$ to the formula for p_n, where $\lfloor x \rfloor$ is the floor function; recall that we set $\lfloor x \rfloor := 0$ if $x < 0$ throughout this paper. Simplifying Assumption 2 provides number-theoretic justification for adding the $\pi_2(p_{n-1}) \sim (n-1)\log\log(n-1) \sim n \log\log n$ term; subtracting the $2\gamma(n-1)$ term

[15] We do not at present know why this deletion produces better accuracy or if it is optimal to this effect.

is an *ad hoc* correction for computational simplicity. Use of the floor function in the correction term ensures our outputs are integer values.

Model 2. The prime numbers $p_1, p_2, p_3, \ldots,$ can be modeled by the sequence having initial value $p_1 = 2$ and for $n \geq 2$ having the values

$$p_n \;=\; 1 \;+\; 2\sum_{k=1}^{n-1}\left\lceil \frac{d(k)}{2} \right\rceil \;+\; \lfloor \pi_2(p_{n-1}) - 2\gamma(n-1) \rfloor\,.$$

Model 2 outputs the same initial values as Model 1 for the estimated sequence of primes, viz. $2, 3, 5, 7, 11, 13, 17, 19, 23, 27$, but provides a better model of prime distribution as n increases. The reader can confirm from Table 1 that Model 2 gives a significantly better numerical approximation to $\pi(n)$ than Model 1 does; in fact, estimates from Model 2 are close to exact at smaller numbers. However, testing Model 2 numerically as n increases becomes increasingly cumbersome: one has to check factorizations for every integer up to p_{n-1}, which is computationally expensive and goes against our goal of having an explicit formulation for the model.[16]

At large numbers, Landau's formula (4.4) approximates $\pi_2(n)$. We will use the relation $\pi_2(p_i) \sim i \log\log i,\ i \geq 2$, by (4.6) to produce a version of Model 2 that is more computationally efficient. Introducing this asymptotic term potentially introduces a source of error, but in practice turns out to yield surprisingly good estimates for $\pi(n)$.

Model 2* (computational version). The prime numbers $p_1, p_2, p_3, \ldots,$ can be modeled by the sequence having initial values $p_1 = 2,\ p_2 = 3$, and for $n \geq 3$ having the values

$$p_n \;=\; 1 \;+\; 2\sum_{k=1}^{n-1}\left\lceil \frac{d(k)}{2} \right\rceil \;+\; \lfloor (n-1)\cdot(\log\log(n-1) - 2\gamma) \rfloor\,.$$

Remark 4.1. Models 2 and 2* can be simplified computationally using equation (2.3).

One can count the number of primes that Model 2* predicts up to any positive integer n. Model 2* still outputs the initial values $2, 3, 5, 7, 11, 13, 17, 19, 23, 27$, and produces an improvement on Model 2 in terms of its approximation to $\pi(n)$ as n increases.

We refer the reader to Table 1 to see that Model 2* yields quite accurate estimates for $\pi(n)$.

[16] We note for clarity that in our computations we use actual primes p_{i-1} inside the Model 2 correction term, not the estimated sequence produced by the model; the Model 2* correction term approximates both cases asymptotically.

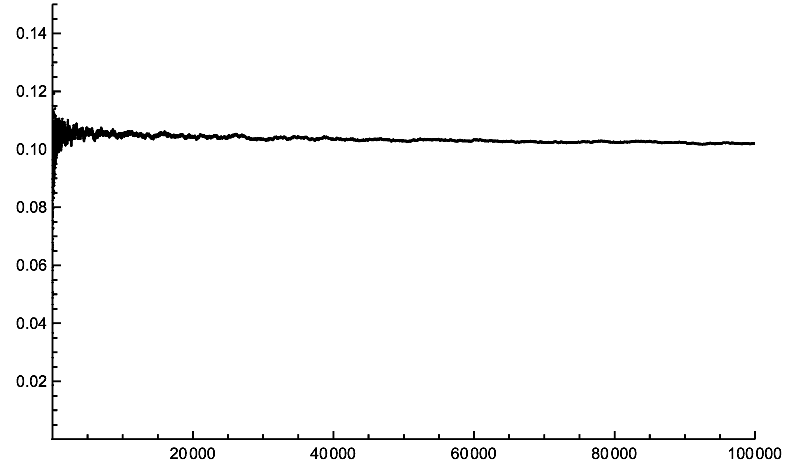

Fig. 2. Relative error in the Model 1 estimate for p_n up to $n = 100{,}000$.

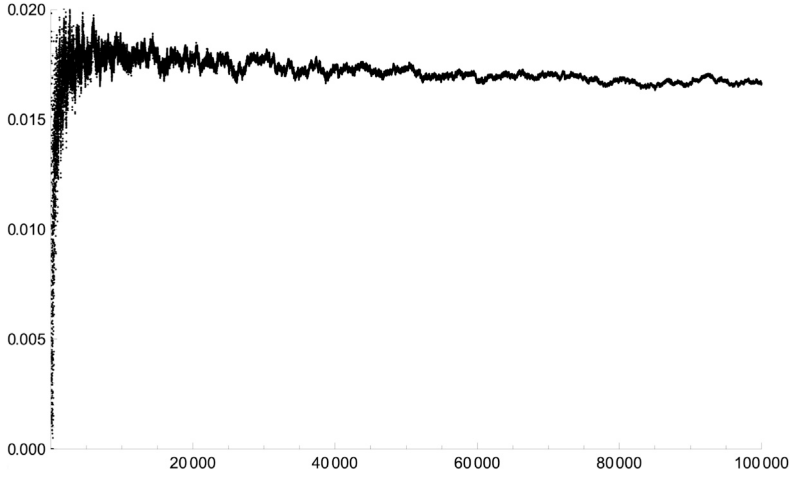

Fig. 3. Relative error in the Model 2 estimate for p_n up to $n = 100{,}000$.

4.2. Comparison of relative errors in the models

From Figs. 2, 3, and 4, one can compare the relative errors in the estimates for p_n provided by Models 1, 2, and 2*, respectively.[17] As the graphs indicate, at $n = 100{,}000$, the relative error in Model 1 is about 0.1, the relative error in Model 2 is about 0.015, and the relative error in Model 2* is below 0.005.

[17] Relative error is computed by $|v_{\text{est}} - v_{\text{act}}|/v_{\text{act}}$ with $v_{\text{est}} \geq 0$ the estimated value and $v_{\text{act}} > 0$ the actual value.

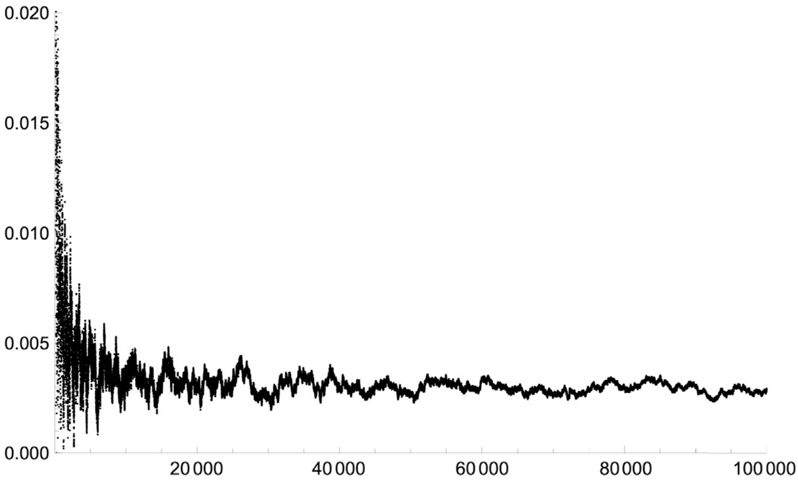

Fig. 4. Relative error in the Model 2* estimate for p_n up to $n = 100{,}000$.

*4.3. Predictions from Model 2 and 2**

With additional correction terms whose necessity largely follows from the premises of Model 1 (but that also include an *ad hoc* adjustment), Models 2 and 2* give surprisingly good estimates for $\pi(n)$ at small numbers. The second-order corrections do not impact the models' compatibility with the prime number theorem asymptotically. As Table 1 displays, Models 2 and 2* are both improvements on Model 1 from a computational perspective.

However, Model 2 requires one to enumerate semiprimes, whereas Model 2* together with the right-hand side of (2.3) provides a fully analytic model of primes requiring no number-theoretic calculations beyond division. Since it involves simpler computations and is evidently more accurate in its estimates of $\pi(n)$ as n increases, we consider Model 2* to be a better model of primes.

We note that the correction terms in Models 2 and 2* *interfere* with one significant prediction of Model 1: eventually, the correction term step functions are strictly increasing as n increases. The strictly increasing correction term prevents the prime gaps in Models 2 and 2* from getting too small as $n \to \infty$.[18] *Models 2 and 2* do not model twin primes at large numbers.*

*4.4. Critique of Models 2 and 2**

We record concerns that might be addressed in the future.

[18] The minimum size of the nth prime gap should grow roughly like $\log\log n$, under Simplifying Assumption 2.

While Models 2 and 2* remain in line with the main term of the prime number theorem, the missing contribution of partitions (and factorizations) of lengths greater than three creeps up, causing the models' estimates to become farther from exact as n increases. Further corrections, either refinements of our assumptions or new ideas entirely, might produce models of higher accuracy.

Recalling that Simplifying Assumption 2 represents an expected overestimate of the prime gap length, then we expect Model 2 produces an *underestimate* for $\pi(n)$; Table 1 supports this as n increases. Furthermore, the approximation of $\pi_2(p_{i-1})$ by Landau's asymptotic (4.4) in Model 2* introduces inaccuracies at values $i \geq 3$ small enough that the asymptotic does not provide a good approximation, and appears to lead to an *overestimate* for $\pi(n)$ in Model 2* by inspection of Table 1. Even so, predictions from Model 2* are generally better than those from Model 2; this appears to be a "happy accident" resulting from the first-order approximation (4.4). Tenenbaum [42] provides an asymptotic series expansion for $\pi_k(n)$; the $k = 2$ case gives

$$\pi_2(n) \;\sim\; \sum_{j=1}^{\infty}(j-1)!\,\frac{n\log\log n}{(\log n)^j} \;+\; \sum_{j=1}^{\infty} C_{j-1}\frac{n}{(\log n)^j} \tag{4.8}$$

as $n \to \infty$, where the C_i are real constants; it is proved in [9] that $C_0 = 0.2614\ldots$ is the *Meissel–Mertens constant.* Substituting $n \mapsto p_n \sim n\log n$, the $j = 1$ summands in (4.8) give

$$\pi_2(p_n) - n\log\log n \;\sim\; C_0 n. \tag{4.9}$$

Thus $(i-1)\log\log(i-1)$ is an *underestimate* for $\pi_2(p_{i-1})$, evidently leading to an overestimate for $\pi(n)$ in Model 2*. Indeed, by inspection of the limited data in Table 1 and consideration of these differing estimates for $\pi(n)$, one can conjecture bounds on $\pi(n)$ as n increases from Models 1, 2, and 2* in the cases shown.

Conjecture 4.1. *Let $\pi^1(n)$ denote the estimate for $\pi(n)$ derived from Model 1, let $\pi^2(n)$ denote the estimate derived from Model 2, and let $\pi^*(n)$ denote the estimate derived from Model 2*. Then as $n \to \infty$ we have*

$$\pi^2(n) \;\leq\; \pi(n) \;\leq\; \pi^*(n) \;\leq\; \pi^1(n).$$

Simplifying Assumptions 1 and 2 suggest $\pi^2(n) \leq \pi(n) \leq \pi^1(n)$. As n increases, we know that $\pi^1(n)$ exceeds $\pi^2(n)$ and $\pi^*(n)$ due to their correction terms, $\pi^2(n) \sim \pi^*(n)$ by (4.4), and $\pi^2(n) \leq \pi^*(n)$ follows from (4.9); we cannot, however, see a reason why $\pi^*(n)$ should always be greater than $\pi(n)$ while $\pi^2(n) \leq \pi(n)$. Table 1 also suggests $n/\log n \leq \pi(n) \leq \mathrm{li}(n)$ as $n \to \infty$, but Littlewood proves $\mathrm{li}(n) - \pi(n)$ changes sign infinitely often [24], so we do not put too much faith in this empirical conjecture.

As we noted in Section 4.3, Models 2 and 2* do not model twin primes at large numbers. To better model small prime gaps, perhaps one can insert a factor $\phi(n) \in [0,1]$

in the correction terms, e.g. $\lfloor \phi(n) \cdot (\pi_2(p_{n-1}) - 2\gamma(n-1)) \rfloor$, such that $\phi(n)$ is closer to zero when $d(n-1)$ is smaller and is closer to one when $d(n-1)$ is greater; for example, $\phi(n) = 1 - 2/d(n-1)$. But this is not assured to produce a better model of prime numbers and may adversely impact estimates for $\pi(n)$.

Finally, we make an aesthetic critique. Model 2 and Model 2* produce sequences of integers that fail to imitate the primes in a key way: the updated models predict *arbitrary primes are sometimes even* as the correction terms can be odd or even. Noting that 2 is the only even prime, we suggest a minor refinement of Models 2 and 2*: replacing $\lfloor x \rfloor$ with a variant $\lfloor x \rfloor^*$ such that $\lfloor x \rfloor^* = \lfloor x \rfloor$ if $\lfloor x \rfloor$ is even, and $\lfloor x \rfloor^* = \lfloor x \rfloor - 1$ if $\lfloor x \rfloor$ is odd. This ensures all modeled prime gaps are even and all predicted prime numbers are odd, and should not significantly affect estimates.

5. Concluding remarks

The authors posit Model 1 as a conceptual model that attempts to explain aspects of prime distribution based on elementary assumptions, which we derive from observations about the partition norm and supernorm. The model not only is compatible with major observations such as the prime number theorem and twin prime conjecture, but also predicts local fluctuations in prime gaps that the prime number theorem does not. We treat these predictions about fluctuations as representing probabilistic statements and give them a cursory check, with mixed findings. By contrast, Model 2 and Model 2* are proposed as computational models that introduce simple corrections to Model 1 to yield estimates for $\pi(n)$ that are surprisingly good.

Because it is compatible with observations of primes and suggestive of new phenomena that can be tested, we hope the approach we present provides a useful complement to probabilistic models.

Acknowledgments

For calculations we use Wolfram Mathematica computer algebra system and Python programming language. The authors are very thankful to Eli DeWitt and Alexander Walker for computer programming that advanced our project, and to Maxwell Schneider for programming consultation. Furthermore, we are grateful to Krishnaswami Alladi, George E. Andrews, William Craig, Philip Cuthbertson, Daniel Desena, Michael Filaseta, Andrew Granville, Timothy Havens, Abhimanyu Kumar, Jeffrey Lagarias, David Leep, Jim McIntyre, Paul Pollack, Ken Ono, Meenakshi Rana, James Sellers, Steven Shawcross, Andrew V. Sills and Hunter Waldron for conversations that informed this paper, and to Ivan Ramirez Zuniga for advice about scientific modeling. In particular, we thank Jeffrey Lagarias and Paul Pollack for offering well-placed suggestions with respect to an earlier draft of this work, Michael Filaseta for invaluable notes on statistical interpretations of our data, and Eli DeWitt for checking our numerical computations prior to publication. We also thank the anonymous referee, who made useful suggestions.

Data availability

Data will be made available on request.

References

[1] K. Alladi, P. Erdős, On an additive arithmetic function, Pac. J. Math. 71 (2) (1977) 275–294.
[2] G.E. Andrews, Partition identities, Adv. Math. 9 (1) (1972) 10–51.
[3] G.E. Andrews, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications, vol. 2, Addison–Wesley, 1976, Reissued: Cambridge University Press, 1998.
[4] T.M. Apostol, Introduction to Analytic Number Theory, Springer Science & Business Media, 1998.
[5] J.E. Cohen, Multiplicative inequalities for primes and the prime counting function, Integers 22 (2022).
[6] S. Corteel, J. Lovejoy, Overpartitions, Trans. Am. Math. Soc. 356 (4) (2004) 1623–1635.
[7] W. Craig, Higher level q-multiple zeta values with applications to quasimodular forms and partitions, arXiv preprint, arXiv:2409.13874, 2024.
[8] W. Craig, J.W. van Ittersum, Ken Ono, Integer partitions detect the primes, Proc. Natl. Acad. Sci. USA 121 (39) (2024) e2409417121.
[9] D. Crişan, R. Erban, On the counting function of semiprimes, Integers 21 (2021).
[10] M.L. Dawsey, M. Just, R. Schneider, A 'supernormal' partition statistic, arXiv preprint, arXiv:2107.14284v1, 2021.
[11] M.L. Dawsey, M. Just, R. Schneider, A 'supernormal' partition statistic, J. Number Theory 241 (2022) 120–141.
[12] T. Forbes, Prime clusters and Cunningham chains, Math. Comput. 68 (228) (1999) 1739–1747.
[13] J.N. Gandhi, Formulae for the Nth prime, in: Proceedings of the Washington State University Conference on Number Theory, 1971, pp. 96–107.
[14] S.W. Golomb, A direct interpretation of Gandhi's formula, Am. Math. Mon. 81 (7) (1974) 752–754.
[15] K. Gomez, MacMahonesque partition functions detect sets related to primes, Arch. Math. 124 (6) (2025) 637–652.
[16] A. Granville, The anatomy of integers and permutations, Preprint, 2008.
[17] A. Granville, K. Soundararajan, Pretentious multiplicative functions and an inequality for the zeta-function, in: CRM Proceedings and Lecture Notes, vol. 46, American Mathematical Society and Centre de Recherches Mathématiques, 2008.
[18] D.R. Heath-Brown, Almost-primes in arithmetic progressions and short intervals, Math. Proc. Camb. Philos. Soc. 83 (3) (1978) 357–375.
[19] M. Just, On a map between partitions and integers (joint work with R. Schneider), in: Modular Forms, Arithmetic and Women in Mathematics Conference (MAAIM), Emory University, Atlanta, Georgia, USA, Nov. 19, 2019.
[20] A. Kumar, Analytic study of norms of prime partitions, Bull. Sci. Math. 171 (2021) 103038.
[21] J.C. Lagarias, Characterizing the supernorm partition statistic, Ramanujan J. 63 (1) (2024) 195–207.
[22] J.C. Lagarias, C. Sun, Asymptotics of reciprocal supernorm partition statistics, Ramanujan J. 65 (2) (2024) 487–520.
[23] E. Landau, Handbuch der Lehre von der Verteilung der Primazahlen, vol. 1, B.G. Teubner, 1909.
[24] J.E. Littlewood, Sur la distribution des nombres premiers, C. R. Acad. Sci. 158 (1914) 1869–1872.
[25] M. Merca, M.D. Schmidt, The partition function $p(n)$ in terms of the classical Möbius function, Ramanujan J. 49 (1) (2019) 87–96.
[26] T.R. Nicely, Detailed instructions for formatting of prime gap record data, https://primegap-list-project.github.io/drtrnicely-format-legacy, Aug. 20, 2019.
[27] K. Ono, The Web of Modularity: Arithmetic of the Coefficients of Modular Forms and q-Series, vol. 102, American Mathematical Society, 2004.
[28] K. Ono, L. Rolen, R. Schneider, Explorations in the theory of partition zeta functions, in: H. Montgomery, A. Nikeghbali, M. Rassias (Eds.), Exploring the Riemann Zeta Function, 190 Years from Riemann's Birth, Springer, 2017, pp. 223–264.
[29] K. Ono, R. Schneider, I. Wagner, Partition-theoretic formulas for arithmetic densities, in: Analytic Number Theory, Modular Forms and q-Series, Springer Proc. Math. Stat. 221 (2017) 611–624.
[30] K. Ono, R. Schneider, I. Wagner, Partition-theoretic formulas for arithmetic densities, II, Hardy-Ramanujan J. 43 (2020) 1–16.

[31] J.B. Rosser, The n-th prime is greater than $n \log n$, Proc. Lond. Math. Soc. 45 (1939) 21–44.
[32] R. Schneider, Partition zeta functions, Res. Number Theory 2 (1) (2016) 9.
[33] R. Schneider, Arithmetic of partitions and the q-bracket operator, Proc. Am. Math. Soc. 145 (5) (2017) 1953–1968.
[34] R. Schneider, Jacobi's triple product, mock theta functions, unimodal sequences and the q-bracket, Int. J. Number Theory 14 (07) (2018) 1961–1981.
[35] R. Schneider, Eulerian series, zeta functions and the arithmetic of partitions, PhD dissertation, Advisor: Ken Ono, Emory University, 2018.
[36] R. Schneider, A.V. Sills, Analysis and combinatorics of partition zeta functions, Int. J. Number Theory (2020) 1–10.
[37] R. Schneider, A.V. Sills, The product of parts or 'norm' of a partition, Integers 20A (2020) A13, 16 pp.
[38] R. Schneider, Partition-theoretic Frobenius-type limit formulas, Ramanujan J. (2024) 1–8.
[39] R. Schneider, Algebra of overpartitions, J. Ramanujan Math. Soc. (2026), in press.
[40] T. Tao, Probabilistic models and heuristics for the primes, https://terrytao.wordpress.com/2015/01/04/254a-supplement-4-probabilistic-models-and-heuristics-for-the-primes-optional, Jan. 4, 2015.
[41] T. Tao, Biases between consecutive primes, https://terrytao.wordpress.com/2016/03/14/biases-between-consecutive-primes, March 14, 2016.
[42] G. Tenenbaum, Introduction to Analytic and Probabilistic Number Theory, vol. 163, American Mathematical Society, 2015.
[43] G. Tenenbaum, M.M. France, P.G. Spain, The Prime Numbers and Their Distribution, vol. 6, American Mathematical Society, 2000.
[44] D. Zagier, Partitions, quasimodular forms, and the Bloch–Okounkov theorem, Ramanujan J. 41 (2016) 345–368.