Many Large Least Primes in Arithmetic Progressions from Short Translates

Shiva Kintali*

August 30, 2026

Abstract

Let $p(k,a)$ denote the least prime congruent to $a \pmod k$, and let $P(k)$ be its maximum over the reduced residue classes. We adapt a reciprocal-logarithmic short-translate weight to dilated offsets and impose a coprimality condition. On one set of moduli of natural density one, this gives

$$P(k) \gg \varphi(k)\frac{\log k(\log_2 k)^2 \log_4 k}{(\log_3 k)^2}.$$

More precisely, for every fixed $0 < \epsilon < 1$, at least $k^{1-\epsilon}$ reduced residue classes have least prime at least c_ϵ times the displayed expression. The proof combines an elementary preliminary cover of progression indices with a finite-interval second-moment argument. It does not require a prime-distribution estimate in a progression whose modulus contains the preliminary sieve modulus. The constructed set includes every sufficiently large modulus with any fixed bound on its number of distinct prime factors.

Keywords. Least prime in an arithmetic progression; sieve methods; prime-free intervals; short translates.

1 Introduction

1.1 The problem and its motivation

All logarithms are natural, and $\log_j$ denotes the j-fold iterated logarithm. Statements containing iterated logarithms are understood for sufficiently large arguments. Implied constants carrying a subscript may depend on the subscripted parameters. We write φ for Euler's totient function and $\pi(v)$ for the number of primes at most v. If $k \geq 2$ and $(a,k) = 1$, let $p(k,a)$ be the least prime congruent to $a \pmod k$, and define

$$P(k) = \max_{\substack{1\leq a<k\\(a,k)=1}} p(k,a).$$

Dirichlet's theorem makes $p(k,a)$ finite. The classical upper-bound problem asks how large this least prime can be in terms of k. Linnik proved that $p(k,a) \ll k^L$ for an absolute constant L [9]; subsequent work reduced the admissible exponent, with Xylouris first obtaining $L = 5.18$ [19] and subsequently showing that $L = 5$ is admissible [20]. Chowla observed that the generalized Riemann hypothesis implies $p(k,a) \ll_\eta k^{2+\eta}$ for every fixed $\eta > 0$, and conjectured the stronger bound

*Email: shiva.kintali@gmail.com Homepage: shivakintali.github.io

$p(k,a) \ll_\eta k^{1+\eta}$ [1]. These upper bounds are important context, but the present paper concerns the complementary extremal question: how large must the worst reduced residue class be?

This lower-bound problem is a progression analogue of constructing long prime-free intervals. The $\varphi(k)$ reduced residue classes are represented by $\varphi(k)$ distinct primes not dividing k, all at most $P(k)$. Hence $\pi(P(k)) \geq \varphi(k)$. Inverting the prime number theorem and using $\log \varphi(k) \sim \log k$ (which follows, for example, from $\varphi(k) \gg k/\log_2 k$) gives

$$P(k) \geq (1+o(1))\varphi(k)\log k.$$

The challenge is to force many consecutive members $a, a+k, \ldots, a+kH$ of one reduced residue class to be composite while retaining quantitative control of both H and the number of admissible choices of a. The factor $\varphi(k)/k$ is the density of integers coprime to k. In our argument it appears both in the count of admissible base points and in the preliminary covering length, where primes dividing k cannot be used.

For later reference, set

$$F(X) = \frac{\log X(\log_2 X)^2 \log_4 X}{(\log_3 X)^2}.$$

A set $\mathcal{K} \subseteq \mathbb{N}$ has natural density one if

$$\lim_{X\to\infty} \frac{1}{X}\#\{k \leq X : k \in \mathcal{K}\} = 1.$$

1.2 Earlier work and the connection with large prime gaps

The lower-bound literature begins with constructions in the spirit of Erdős and Rankin for long strings of composite integers [2, 13]. Let $p'(k,\ell)$ denote the least prime greater than k that is congruent to $\ell \pmod k$, where $k \geq 2$ and $(k,\ell) = 1$. For each fixed positive integer ℓ, Prachar [12] and Schinzel [14] proved, along an infinite sequence of moduli coprime to ℓ, the lower bound

$$p'(k,\ell) \gg k \log k \log_2 k \frac{\log_4 k}{(\log_3 k)^2}.$$

Wagstaff obtained an analogous result when the modulus is prime [17]. A modification of the argument of Hensley and Richards [8] also shows that $P(k)/(\varphi(k)\log k)$ is unbounded through prime moduli; see [10, Section 1].

Pomerance first strengthened the elementary baseline to $P(k) \geq (e^{\gamma_{\mathrm{E}}} + o(1))\varphi(k)\log k$ as $k \to \infty$ through all integers, where γ_{E} is Euler's constant [11, Theorem 2]. He also established a density-one result of a different strength [11]: for almost all k,

$$P(k) \geq (e^{\gamma_{\mathrm{E}}} + o(1))\varphi(k)\log k \log_2 k \frac{\log_4 k}{(\log_3 k)^2}. \tag{1}$$

Granville and Pomerance subsequently obtained, for infinitely many coprime pairs (k,ℓ), the estimate [7, Theorem 1]

$$p(k,\ell) \geq (2e^{\gamma_{\mathrm{E}}} + o(1))k \log k \log_2 k \frac{\log_4 k}{(\log_3 k)^2}.$$

This is an explicit leading-constant improvement at the displayed k-scale along a selected sequence; because it is expressed in terms of k, rather than $\varphi(k)$, and has an infinitely-many-moduli quantifier, it is not directly comparable with a density-one theorem for $P(k)$. These results make clear that the arithmetic of the modulus and the efficiency of the underlying residue-class cover have to be controlled simultaneously.

The modern large-gap construction of Ford, Green, Konyagin, Maynard, and Tao (FGKMT) combines a multidimensional prime-detecting sieve with a quantitative hypergraph covering argument. Writing p_n for the nth prime, their result is [3]

$$\max_{p_{n+1}\le X}(p_{n+1}-p_n)\gg \frac{\log X\log_2 X\log_4 X}{\log_3 X}.$$

Li, Pratt, and Shakan imported this circle of ideas into the least-prime problem. Their principal new device is a sieve weight that detects primes together with small multiples of primes. Write $\omega(k)$ for the number of distinct prime divisors of k. For every fixed $0<\eta<1/2$, Li–Pratt–Shakan proved [10, Theorem 1.1] that every sufficiently large k satisfying

$$\omega(k)\le \exp\!\left(\left(\frac{1}{2}-\eta\right)\frac{\log_2 k\,\log_4 k}{\log_3 k}\right)$$

obeys the effective estimate

$$P(k)\gg_\eta \varphi(k)\log k\log_2 k\frac{\log_4 k}{\log_3 k}. \tag{2}$$

For each such η, the displayed restriction on $\omega(k)$ holds on a set of natural density one. Indeed, its right-hand side eventually exceeds $(\log_2 k)^2$, whereas $\sum_{k\le X}\omega(k)\ll X\log_2 X$. Splitting off $k\le\sqrt{X}$ and applying Markov's inequality to the remaining integers bounds the number of exceptions up to X by $O(\sqrt{X}+X/\log_2 X)=o(X)$.

In a preliminary preprint primarily concerned with simultaneously small and large consecutive-prime gaps, Sun and Pan state a fixed-positive-residue theorem for $p'(k,\ell)$ [15, Theorem 6.2]. Its least-prime component has the classical Erdős–Rankin iterated-logarithm scale and an infinitely-many-moduli quantifier; it is not an input to our proof and does not supersede (2). Wagstaff proposed a probabilistic model for the size of $P(k)$ [18]. Fiori subsequently refined that model, conjectured in particular that

$$P(k)<3\varphi(k)\log k\log\varphi(k)\qquad(k>3),$$

and verified this inequality computationally through $k\le 10^8$ [4]. These probabilistic and computational conclusions are useful benchmarks but are not unconditional asymptotic bounds.

Two unpublished companion manuscripts provide the immediate technical background for this paper. The tilted residue law of [16] separates a preliminary random sieve from a later completion of its exceptional set. The short-translate construction of [5, Proposition 1.2 and Section 4] replaces the prime-detecting layer by reciprocal-logarithmic divisor coefficients and a finite-interval second moment. The latter mechanism is the one adapted here. Unlike the FGKMT–Li–Pratt–Shakan route, the proof below does not need a positive lower bound for primes under a Maynard-type weight or a hypergraph covering theorem.

1.3 Main result

Our theorem strengthens (2) on a density-one set and gives many witnessing residue classes, not merely the largest one.

Theorem 1.1. *There is a set $\mathcal{K}$ of natural density one such that, for every fixed $0<\epsilon<1$, there is a constant $c_\epsilon>0$ for which every sufficiently large $k\in\mathcal{K}$ satisfies*

$$\#\{1\le a<k:(a,k)=1,\ p(k,a)>c_\epsilon\varphi(k)F(k)\}\ge k^{1-\epsilon}.$$

In particular, there is an absolute constant $c>0$ such that every sufficiently large $k\in\mathcal{K}$ satisfies $P(k)\ge c\varphi(k)F(k)$.

The proof in Section 4 takes

$$\mathcal{K} = \left\{ k \geq k_0 : \prod_{p|k} (1 - p^{-3/4})^{-1} \leq \tfrac{1}{2} \log_3 k \right\},$$

where k_0 is a sufficiently large fixed integer. This explicit choice gives a further consequence: for every fixed integer $r \geq 1$, the conclusions hold for all sufficiently large k with $\omega(k) \leq r$. Indeed, for such k,

$$\prod_{p|k} (1 - p^{-3/4})^{-1} \leq (1 - 2^{-3/4})^{-r},$$

whereas $\frac{1}{2} \log_3 k \to \infty$. In particular, the theorem applies to all sufficiently large prime moduli and prime-power moduli. The threshold here may depend on r and ϵ. This is a consequence of the explicit set, not of natural density one alone.

The ratio of the new scale $\varphi(k)F(k)$ to the Li–Pratt–Shakan scale in (2) is

$$\frac{\log_2 k}{\log_3 k} \longrightarrow \infty.$$

Relative to the density-one bound (1), the gain is a factor of order $\log_2 k$. The multiplicity assertion is stronger than an estimate for $P(k)$: for every fixed $0 < \epsilon < 1$, at least $k^{1-\epsilon}$ distinct reduced classes have no prime below the stated scale. The density-one set $\mathcal{K}$ is independent of ϵ.

1.4 Comparison with previous bounds and conjectural scale

To compare results with the same normalization, put

$$R(k) = \frac{P(k)}{\varphi(k) \log k}.$$

The prime-number-theorem argument gives $R(k) \geq 1 + o(1)$, and Pomerance's all-moduli theorem strengthens this to $R(k) \geq e^{\gamma_{\mathrm{E}}} + o(1)$. His density-one theorem (1) gives

$$R(k) \geq (e^{\gamma_{\mathrm{E}}} + o(1)) \frac{\log_2 k \, \log_4 k}{(\log_3 k)^2},$$

whereas Li–Pratt–Shakan give, on a density-one family,

$$R(k) \gg \frac{\log_2 k \, \log_4 k}{\log_3 k}.$$

Theorem 1.1 gives

$$R(k) \gg \frac{(\log_2 k)^2 \log_4 k}{(\log_3 k)^2}.$$

Consequently, at the level of directly comparable density-one lower bounds, the gain over Pomerance is a factor of order $\log_2 k$, and the gain over Li–Pratt–Shakan is

$$\frac{\log_2 k}{\log_3 k} \longrightarrow \infty.$$

The improvement is therefore one of asymptotic order and not merely of the implied constant. On the other hand, Pomerance records the leading constant $e^{\gamma_{\mathrm{E}}}$ at his scale, and Granville–Pomerance

obtain the leading factor $2e^{\gamma_{\mathrm{E}}} + o(1)$ on their selected sequence, while the present theorem makes no assertion about an optimal leading constant.

The quantifiers must be kept separate. The results of Prachar and Schinzel concern a fixed residue ℓ along an infinite sequence of moduli; Wagstaff gives an analogous result for prime moduli. The modification of the Hensley–Richards argument concerns the maximum $P(k)$ through prime moduli. Although a density-one set need not contain any primes, the explicit choice of $\mathcal{K}$ above shows that our bound holds for every sufficiently large prime modulus, and more generally whenever $\omega(k)$ is bounded. Our theorem supplies many classes but does not prescribe a fixed residue ℓ. Nor do we obtain our bound for every modulus satisfying the Li–Pratt–Shakan restriction on $\omega(k)$; our improvement is on a density-one set. Likewise, the FGKMT theorem is a theorem about gaps between consecutive primes and small-prime covers; the present theorem is a theorem about least primes in residue classes and is not a strengthening of the FGKMT gap bound.

Granville–Pomerance also give a quantitative count of progressions with large least prime, aggregated over moduli up to a given bound [7, Theorem 1]. The multiplicity assertion here has a different quantifier: Theorem 1.1 supplies at least $k^{1-\epsilon}$ distinct reduced residue classes at the new scale for *each* sufficiently large k in one density-one set. This is stronger than an estimate for the maximum $P(k)$ alone, but should not be confused with the earlier aggregate count across moduli.

Probabilistic models predict a substantially larger ultimate order. The coupon-collector model of Li–Pratt–Shakan [10, Section 2] suggests

$$\liminf_{k\to\infty} \frac{P(k)}{\varphi(k)(\log k)^2} = 1, \qquad \limsup_{k\to\infty} \frac{P(k)}{\varphi(k)(\log k)^2} = 2,$$

while the Wagstaff–Fiori model supports the comparable scale $\varphi(k)\log k \log\varphi(k)$ [18, 4]. The ratio of the scale in Theorem 1.1 to $\varphi(k)(\log k)^2$ is

$$\frac{(\log_2 k)^2 \log_4 k}{\log k\,(\log_3 k)^2} \longrightarrow 0.$$

Thus the theorem gives an iterated-logarithmic improvement over the cited published density-one lower bounds, but it remains far below the conjectural order.

For clarity, Theorem 1.1 does not establish the stronger estimate

$$P(k) \gg \varphi(k)\log k \frac{(\log_2 k)^2}{\log_3 k}.$$

Indeed, its scale divided by the last display equals $\log_4 k/\log_3 k \to 0$. Equivalently, the stronger expression is larger by the unbounded factor $\log_3 k/\log_4 k$. We include this comparison to delimit precisely what is and is not proved here.

The theorem should not be read as a stronger small-prime covering theorem of FGKMT type. The surviving prime indices in the preliminary sieve are counted by an upper bound; the short-translate weight itself is not required to give a positive lower bound for primes. It establishes compositeness and does not by itself force any prescribed progression term to be prime. In the completion step below, the exceptional progression terms may receive proper prime factors larger than the preliminary sieve level x. Thus the conclusion controls the first prime in many residue classes, but the argument neither covers the whole index interval using only primes at most x nor shows that the auxiliary primes can be adjoined to the preliminary sieve modulus at logarithmic cost $O(x)$.

1.5 What is novel and what is derived

What is novel relative to the cited inputs. The contribution established here is the density-one lower bound at the scale

$$\varphi(k)\frac{\log k(\log_2 k)^2\log_4 k}{(\log_3 k)^2},$$

together with the $k^{1-\epsilon}$ multiplicity statement. The following ingredients are proved in this paper rather than invoked as consequences of earlier work.

- The short-translate weight is made uniform for the dilated forms $b+Mt+kj$ after every auxiliary prime dividing kM is deleted. For an auxiliary prime $p \nmid kM$, two local roots attached to $j \neq j'$ coincide if and only if $p \mid j-j'$; thus $p > |j-j'|$ suffices for separation, independently of the dilation factor k.
- The reduced-residue constraint $(b+Mt,k)=1$ is imposed inside the weighted finite-interval count by inclusion–exclusion. The proof retains a quantitative supply of at least $\rho(k)T/(2D^3)$ successful translates, where $\rho(k)=\varphi(k)/k$.
- A density-one estimate for the forbidden prime divisors of k is coupled to a preliminary covering lemma in which those primes are removed. The remaining exceptional set has size at most δx uniformly over the permitted moduli.
- The final Chinese-remainder transfer is arranged so that $M(T+1)<k$. Consequently distinct successful translates give distinct representatives $1 \le a < k$, which yields the stated multiplicity as well as the lower bound for $P(k)$.

What is derived. The reciprocal-logarithmic coefficients, the restriction to pairwise coprime divisor tuples, and the shared product cutoff are adapted from the short-translate argument of [5]. The conceptual division into a preliminary sieve and an exception-completion stage is motivated by [16]. The coefficient estimates use the uniform sieve summation theorem [6, Theorem 4.4], together with the prime number theorem and Mertens' estimates; Dirichlet's theorem is used to ensure that $p(k,a)$ is defined. The historical Erdős–Rankin covering idea, the modern FGKMT sieve framework, and the Li–Pratt–Shakan treatment of least primes explain the surrounding strategy [13, 3, 10], but neither the FGKMT hypergraph theorem nor the Li–Pratt–Shakan prime-detecting weight is used as a black box in the proof. Likewise, the global prime-gap conclusions of [16, 5] are not assumed: every extension of the short-translate mechanism needed for the present arithmetic-progression setting is established in Sections 2–4. In particular, the preliminary covering theorem of [16] is not used as a black box.

1.6 Proof strategy

We briefly describe how the pieces fit together. Let k lie in a dyadic interval $(K,2K]$ and take x to be a small constant multiple of $\log K$. The proof has four stages.

First, Section 2 constructs normalized reciprocal-logarithmic coefficients supported on squarefree products of primes in $(x^C, \exp(x/(\log x)^5)]$, after primes dividing a prescribed integer have been deleted. Ford's summation theorem supplies estimates that are uniform in the deleted set.

Second, these coefficients are inserted into a weight for the linear forms $b+Mt+kj$, with t restricted to a finite interval. Expanding the second moment, exploiting the pairwise-coprime support, and controlling the shared product cutoff show that many t give every exceptional form a

proper auxiliary prime factor. Inclusion–exclusion over the divisors of k simultaneously enforces $(b + Mt, k) = 1$.

Third, Section 3 introduces

$$V(k) = \prod_{p \mid k} (1 - p^{-3/4})^{-1}.$$

This quantity has bounded mean, so $V(k)$ is small for a density-one set of moduli. For such k, an elementary Rankin argument and a greedy residue choice cover all but $O(x)$ of the indices in an interval of length

$$H \asymp \rho(k) x \frac{(\log x)^2 \log_3 x}{(\log_2 x)^2}.$$

Finally, Section 4 chooses b by the Chinese remainder theorem. The preliminary residues make the covered forms composite for every t, and the short-translate lemma completes the exceptional forms for many t. Parameter choices ensure $1 \le b + Mt < k$, while $kH \gg_\epsilon \varphi(k)F(k)$. Counting the successful translates gives at least $k^{1-\epsilon}$ different reduced classes.

1.7 Organization of the paper

Section 2 proves the coefficient estimates and the conditioned short-translate lemma. Section 3 establishes the density-one control of forbidden primes and constructs the preliminary cover. Section 4 combines the two ingredients, performs the Chinese-remainder transfer, and proves Theorem 1.1.

2 A uniform supply of reduced-residue short translates

The first lemma supplies coefficients whose full sum cancels and whose quadratic mass is controlled. The second uses these coefficients to give every prescribed form an auxiliary prime divisor, while retaining a coprimality condition on its base point. The latter argument counts translates in a finite interval; it does not merely produce a residue class modulo the auxiliary product.

We use μ for the Möbius function, $\omega(n)$ for the number of distinct prime divisors of n, and $\mathrm{rad}(n)$ for their product. Empty products are 1, so $\omega(1) = 0$ and $\mathrm{rad}(1) = 1$. All asymptotic estimates in this section are as $x \to \infty$, uniformly in the integer data satisfying the stated hypotheses; constants and thresholds may depend on the fixed parameters.

Lemma 2.1 (Coefficient estimates after deleting forbidden primes)**.** *Fix constants $C > 6$ and $C_0 > 0$. Let x tend to infinity and put*

$$z = x^C, \qquad y = \exp\left(\frac{x}{(\log x)^5}\right), \qquad \beta = \frac{1}{\log y}.$$

For a positive integer N with $\log N \le C_0 x$, define

$$P = \prod_{\substack{z < p \le y \\ p \nmid N}} p, \qquad B = \sum_{\substack{d \mid P \\ d > 1}} \frac{1}{\varphi(d) \log d}.$$

For all sufficiently large x, one has $B \asymp 1$ uniformly in N. For such x, set

$$a(1) = 1, \qquad a(d) = -\frac{1}{B \log d} \quad (d \mid P,\ d > 1).$$

Then, uniformly in N,

$$\sum_{d\mid P}\frac{a(d)}{\varphi(d)}=0,\qquad |a(d)|\le 1,\qquad A:=\sum_{d\mid P}\frac{a(d)^2}{\varphi(d)}=1+O((\log x)^{-2}), \tag{3}$$

$$\sum_{\substack{d\mid P\\ d>1}}\frac{|a(d)|d^{\gamma}}{\varphi(d)}\ll 1,\qquad A_\gamma:=\sum_{d\mid P}\frac{a(d)^2d^{\gamma}}{\varphi(d)}\le A+C_1\gamma\quad(0\le\gamma\le 2\beta), \tag{4}$$

$$\sum_{\substack{d\mid P\\ p\mid d}}\frac{|a(d)|}{\varphi(d)}\ll\frac{1}{p}\quad(p\mid P). \tag{5}$$

The constants may depend on C and C_0.

Proof. First take $P_0=\prod_{z<p\le y}p$, without deleting primes. Write

$$W(z)=\prod_{p\le z}(1-1/p),\qquad M_z(v)=\sum_{\substack{d\le v\\ (d,\prod_{p\le z}p)=1}}\frac{\mu(d)^2}{\varphi(d)}.$$

Ford's sieve summation theorem [6, Theorem 4.4] implies

$$M_z(v)=W(z)\log v+O(1)\qquad(z,v\ge 2), \tag{6}$$

with an absolute implied constant. To check the hypotheses, let g be the multiplicative function supported on squarefree integers with $g(p)=\mathbf{1}_{p>z}/p$. Then $0\le g(p)\le 1/2$. Mertens' estimate $\sum_{p\le v}(\log p)/p=\log v+O(1)$ gives, for an absolute $K_0>0$,

$$-\log z-K_0\le\sum_{w\le p\le v}g(p)\log p-\log(v/w)\le K_0\qquad(2\le w\le v).$$

Thus Ford's parameters can be taken as dimension 1, $A_1=1/2$, $A_2=K_0$, and loss $L=\log z+K_0$. The associated multiplicative function h has $h(p)=g(p)/(1-g(p))=\mathbf{1}_{p>z}/(p-1)$, and the singular series is $W(z)$. Both error terms in the theorem are $O(W(z)(L+1))=O(1)$, proving (6).

Put $B_0=\sum_{d\mid P_0,d>1}(\varphi(d)\log d)^{-1}$ and $U_\gamma=\sum_{d\mid P_0}d^{\gamma}/\varphi(d)$. Mertens' theorem gives $U_0\asymp\log y/\log z$. For $0\le\gamma\le 2\beta$,

$$\frac{U_\gamma}{U_0}=\prod_{z<p\le y}\left(1+\frac{p^{\gamma}-1}{p}\right)\le\exp\left(C_2\gamma\sum_{p\le y}\frac{\log p}{p}\right)\ll 1,$$

because $\gamma\log p\le 2$. The part of B_0 with $d>y$ is at most $U_0/\log y\ll 1/\log z$. For $d\le y$, the squarefree integers coprime to $\prod_{p\le z}p$ are exactly the divisors of P_0 in this range. Since $M_z(z)=1$, partial summation gives

$$\sum_{\substack{d\mid P_0\\ 1<d\le y}}\frac{1}{\varphi(d)\log d}=\frac{M_z(y)-1}{\log y}+\int_z^y\frac{M_z(v)-1}{v(\log v)^2}\,dv.$$

Substituting (6) and adding the tail consequently gives

$$B_0=W(z)\log\frac{\log y}{\log z}+O\Big(\frac{1}{\log z}\Big)\longrightarrow\frac{e^{-\gamma_{\mathrm{E}}}}{C}>0, \tag{7}$$

where γ_{E} is Euler's constant: indeed, $W(z) \sim e^{-\gamma_{\mathrm{E}}}/(C \log x)$ and $\log(\log y/\log z) = \log x - 6\log\log x - \log C$. For the quadratic coefficient sum, first enlarge to all squarefree integers whose prime factors exceed z. A second partial summation gives

$$\sum_{\substack{d \mid P_0 \\ d>1}} \frac{1}{\varphi(d)(\log d)^2} \le 2\int_z^\infty \frac{M_z(v)-1}{v(\log v)^3}\,dv \ll (\log z)^{-2}. \tag{8}$$

The boundary term at infinity vanishes because $M_z(v) = O(W(z)\log v + 1)$. The integral is $O(W(z)/\log z + (\log z)^{-2}) = O((\log z)^{-2})$.

Let $E = \{p : z < p \le y,\ p \mid N\}$. Since $\log N \le C_0 x$,

$$\sum_{p\in E} \frac{1}{p} \le \frac{\log N}{z\log z} \ll \frac{x}{z\log z} = o(1).$$

Writing $d = pe$ shows, for $p \mid P_0$,

$$\sum_{\substack{d \mid P_0 \\ p \mid d}} \frac{1}{\varphi(d)\log d} \le \frac{1}{p-1}\left(\frac{1}{\log p} + B_0\right).$$

Summing over the deleted primes (allowing overcounting) gives $0 \le B_0 - B \ll \sum_{p\in E} 1/p = o(1)$, proving $B \asymp 1$ and in particular $B > 0$ for all sufficiently large x. The cancellation identity follows from the definition of B. For $d > 1$, the bound $|a(d)| \le 1$ follows from $B\log d \ge B\log z \to \infty$; for $d = 1$ it is immediate. Equation (8) proves the assertion about A.

For the first estimate of (4), split the sum at $d = y$. Below y, use $d^\gamma \le e^2$ and the identity

$$\sum_{\substack{d \mid P \\ d>1}} \frac{|a(d)|}{\varphi(d)} = 1.$$

Above y use

$$\sum_{\substack{d \mid P \\ d>y}} \frac{|a(d)|d^\gamma}{\varphi(d)} \le \frac{U_\gamma}{B\log y} \ll 1.$$

The inequality $d^\gamma - 1 \le \gamma d^\gamma \log d$ now yields

$$A_\gamma - A \le \frac{\gamma}{B}\sum_{\substack{d \mid P \\ d>1}} \frac{|a(d)|d^\gamma}{\varphi(d)} \ll \gamma.$$

Finally, separating $d = p$ and then writing $d = pe$ gives

$$\sum_{\substack{d \mid P \\ p \mid d}} \frac{|a(d)|}{\varphi(d)} \le \frac{1}{B(p-1)}\left(\frac{1}{\log p} + B\right) \ll \frac{1}{p}.$$

□

Lemma 2.2 (Short translates in reduced residue classes)**.** *Fix $A_0 > 1$, $C_0 > 0$, and $\tau > 0$. Choose a fixed $C > \max(A_0, 6)$ and $0 < \kappa < \tau/3$. There exists $\delta > 0$ such that the following holds for all*

sufficiently large x. Let k, M be coprime positive integers with $\log(kM) \le C_0 x$, let $b \in \mathbb{Z}$, and let J be a set of at most δx distinct integers contained in an interval of length x^{A_0}. Put

$$T = \lfloor e^{\tau x} \rfloor, \quad D = e^{\kappa x}, \quad z = x^C, \quad y = \exp\left(\frac{x}{(\log x)^5}\right).$$

Also put

$$P = \prod_{\substack{z < p \le y \\ p \nmid kM}} p, \quad \rho = \frac{\varphi(k)}{k}.$$

In any interval I consisting of T consecutive integers, at least $\rho T/(2D^3)$ integers t satisfy

$$(b + Mt, k) = 1, \qquad (b + Mt + kj, P) > 1 \quad \text{for every } j \in J. \tag{9}$$

For each successful t, if $b + Mt + kj > y$ for every $j \in J$, then all these values are composite. The constant δ and the threshold for x depend only on $A_0, C_0, \tau, C, \kappa$; in particular they are independent of b and the locations of I and J.

Proof. Let $C_1 \ge 1$ be a constant valid in (4), and fix

$$0 < \delta \le \min\left(1, \frac{\kappa}{4C_1}\right).$$

Throughout the proof put $\beta = 1/\log y$ and $r = |J|$. We first record uniform bounds for the coprimality condition:

$$\omega(k) = O(x/\log x), \qquad \rho^{-1} \le 2^{\omega(k)} = \exp(O(x/\log x)). \tag{10}$$

For the first bound, if $m = \omega(k)$, then $k \ge m!$; Stirling's inequality and $\log k \le C_0 x$ give the result. The second follows by bounding each factor $p/(p-1)$ by 2. If $r = 0$, inclusion–exclusion and $(M, k) = 1$ give

$$\#\{t \in I : (b + Mt, k) = 1\} = \rho T + O(2^{\omega(k)}) \ge \tfrac{1}{2}\rho T \ge \frac{\rho T}{2D^3}$$

for large x, by (10). Thus assume $r \ge 1$ and enumerate $J = \{j_1, \dots, j_r\}$.

Step 1: local moments and a quadratic-form estimate. Put $L_i(t) = b + Mt + kj_i$. For $p \mid P$, define

$$\psi_{i,p}(t) = \frac{1 - p\mathbf{1}_{p \mid L_i(t)}}{p - 1}, \qquad \psi_{i,d}(t) = \prod_{p \mid d} \psi_{i,p}(t), \qquad \psi_{i,1}(t) = 1.$$

The coefficient M is invertible modulo p. Moreover the roots of L_i and L_j are distinct when $i \ne j$: equality of the roots would imply $p \mid k(j_i - j_j)$, whereas $p \nmid k$ and $0 < |j_i - j_j| \le x^{A_0} < p$. Writing $\mathbb{E}_p$ for the average over $t \pmod p$, the complete local averages are therefore

$$\mathbb{E}_p \psi_{i,p} = 0, \qquad \mathbb{E}_p \psi_{i,p}^2 = \frac{1}{p-1}, \qquad \mathbb{E}_p \psi_{i,p}\psi_{j,p} = -\frac{1}{(p-1)^2} \quad (i \ne j). \tag{11}$$

Use the coefficients from Lemma 2.1 with $N = kM$. For $0 \le s \le r$, let $\mathcal{R}_s$ be the ordered s-tuples of pairwise coprime divisors of P whose product is at most D; the empty tuple is the sole element of $\mathcal{R}_0$. Define

$$R(t) = \sum_{\boldsymbol{d} \in \mathcal{R}_r} \prod_{i=1}^{r} a(d_i)\psi_{i,d_i}(t).$$

All complete expectations $\mathbb{E}$ below are over $t \pmod P$. The Chinese remainder theorem makes these local coordinates independent.

We first give a quadratic-form estimate that will also be used with different coefficients. If two tuples in $\mathcal{R}_s$ have different total products, their associated functions are orthogonal by the first identity in (11). For tuples with a common product m, the diagonal norm squared is $1/\varphi(m)$; multiply every associated function by $\sqrt{\varphi(m)}$ to give it squared L^2 norm one. Keeping a prime at the same coordinate gives inner product 1; changing its coordinate gives $-1/(p-1)$. For $s \ge 1$ and fixed m, the tuples correspond exactly to assignments of each prime factor of m to one of the s coordinates. All assignments are retained because their total product is the same $m \le D$. The local choices therefore factor, and each absolute off-diagonal row sum is bounded by

$$\prod_{p\mid m}\left(1+\frac{s-1}{p-1}\right)-1 \ll \frac{r\log D}{z} = o(1).$$

Indeed, for $s \ge 1$ the product is at most $\exp(2r\omega(m)/z)$, and $\omega(m) \le \log D/\log z$. For $s = 0$ there is only the empty tuple, so the off-diagonal sum is zero. Applying $2|uv| \le u^2+v^2$ to the off-diagonal terms proves, for arbitrary real coefficients $c_{\boldsymbol{d}}$,

$$\mathbb{E}\left(\sum_{\boldsymbol{d}\in\mathcal{R}_s} c_{\boldsymbol{d}}\prod_{i=1}^{s}\psi_{i,d_i}\right)^2 = \left(1+O\Big(\frac{r\log D}{z}\Big)\right)\sum_{\boldsymbol{d}\in\mathcal{R}_s}\frac{c_{\boldsymbol{d}}^2}{\prod_i\varphi(d_i)}. \tag{12}$$

Step 2: losses from the product cutoff and repeated primes. Put $u = \log D/\log y = \kappa(\log x)^5$. Summing over all ordered r-tuples of divisors of P, without either restriction, gives diagonal mass A^r. Rankin's inequality and (4) bound the part beyond the product cutoff by

$$D^{-\beta}A_\beta^r \le A^r\exp(-u+C_1r\beta) \le A^re^{-u/2},$$

because $C_1r\beta \le (C_1\delta/\kappa)u \le u/4$. The mass with a prime common to two coordinates is at most

$$\binom{r}{2}A^{r-2}\sum_{p\mid P}\left(\sum_{\substack{d\mid P\\ p\mid d}}\frac{a(d)^2}{\varphi(d)}\right)^2 \ll A^r\frac{r^2}{z} = o(A^r).$$

For $r = 1$ this loss is zero. The bound uses $|a(d)| \le 1$, (5), and $\sum_{n>z} n^{-2} \ll 1/z$. Hence (12) yields

$$\mathbb{E}R^2 = (1+o(1))A^r. \tag{13}$$

All losses here are relative to A^r; no assertion that $A^r = 1+o(1)$ is needed or made.

Step 3: controlling uncovered forms. For a fixed coordinate i, replace each factor $\psi_{i,d}$ in R by $1/\varphi(d)$ and call the resulting function R_i. On the event $(L_i(t), P) = 1$, this changes no value. Thus, pointwise,

$$R(t)^2\mathbf{1}_{(L_i(t),P)=1} \le R_i(t)^2. \tag{14}$$

After relabelling so that $i = r$, the coefficient attached to the other coordinates with product m is

$$\epsilon(m)\prod_{j<r}a(d_j), \qquad \epsilon(m) = \sum_{\substack{d\mid P\\ d\le D/m\\ (d,m)=1}}\frac{a(d)}{\varphi(d)}.$$

Since $m \le D$, the summand $d = 1$ is retained. The full coefficient sum is zero and every summand with $d > 1$ is negative. Consequently $\epsilon(m)$ is nonnegative and is bounded by the absolute mass of the excluded terms:

$$0 \le \epsilon(m) \le \sum_{\substack{d \mid P \\ d > D/m}} \frac{|a(d)|}{\varphi(d)} + \sum_{p \mid m} \sum_{\substack{d \mid P \\ p \mid d}} \frac{|a(d)|}{\varphi(d)} \ll (m/D)^{\beta} + \frac{\log D}{z}.$$

For the tail we used $\mathbf{1}_{d > D/m} \le (dm/D)^{\beta}$ and (4); here $D/m \ge 1$, so the tail excludes $d = 1$. For the other term, (5) gives $\sum_{p \mid m} p^{-1} \le \omega(m)/z \le \log D/(z \log z)$. Use (12) with $s = r - 1$, square this bound, and enlarge the nonnegative diagonal sum by dropping its restrictions. This gives

$$\mathbb{E} R_i^2 \ll D^{-2\beta} A_{2\beta}^{r-1} + \frac{(\log D)^2}{z^2} A^{r-1} \ll A^r \left(e^{-u} + \frac{(\log D)^2}{z^2} \right).$$

The last inequality follows from

$$\frac{D^{-2\beta} A_{2\beta}^{r-1}}{A^{r-1}} \le \exp(-2u + 2C_1(r-1)\beta) \le e^{-u},$$

by the fixed choice of δ. This also covers $r = 1$, using $A_{2\beta}^0 = 1$ and the empty-tuple case of (12). In particular,

$$\sum_{i=1}^{r} \mathbb{E} R_i^2 \ll r A^r \left(e^{-u} + \frac{(\log D)^2}{z^2} \right) = o(A^r). \tag{15}$$

Step 4: transferring the moments to the finite interval. The exact divisor expansion is

$$\psi_{i,d}(t) = \frac{1}{\varphi(d)} \sum_{e \mid d} \mu(e) e \, \mathbf{1}_{e \mid L_i(t)}.$$

Expand R by this identity, keeping the resulting terms uncollected. The sum of their absolute coefficients is at most

$$\sum_{\substack{n \mid P \\ n \le D}} r^{\omega(n)} \frac{\sigma_1(n)}{\varphi(n)} \le 2 D^{1 + \log r / \log z} \le D^{3/2}. \tag{16}$$

Here $\sigma_1(n) = \sum_{d \mid n} d$. Indeed, the number of assignments of the primes of n to the r coordinates is $r^{\omega(n)}$, each original coefficient has absolute value at most one, and expanding the local factors contributes $\prod_{p \mid n} (p+1)/(p-1) = \sigma_1(n)/\varphi(n)$. This last product is at most 2 for large x, since $\omega(n) \le \log D / \log z$ and $p > z$. Also $\log r / \log z \le 1/C + o(1) < 1/6 + o(1)$, which proves the last inequality including the factor 2. Each R_i has the same bound: in the indicator expansion, the substitution defining R_i keeps only $e = 1$ in that coordinate. It therefore deletes terms from the uncollected expansion and cannot increase this sum of absolute coefficients.

Put $\chi(t) = \mathbf{1}_{(b+Mt,k)=1}$. Since $(M, k) = 1$, inclusion-exclusion gives

$$\chi(t) = \sum_{d \mid \mathrm{rad}(k)} \mu(d) \mathbf{1}_{d \mid b + Mt}.$$

A compatible auxiliary divisibility system specifies one residue class modulo some $q \mid P$. As $(q, k) = 1$, adjoining $d \mid b + Mt$ specifies one residue modulo dq. The average of its indicator over I differs from $1/(dq)$ by at most $1/T$. This elementary interval-counting bound holds even when $dq > T$; no

equidistribution estimate for primes is involved. For each compatible system, inclusion–exclusion has main term

$$\sum_{d|\mathrm{rad}(k)} \frac{\mu(d)}{dq} = \frac{\rho}{q}.$$

The coefficient mass after squaring is at most D^3, and inclusion–exclusion has $2^{\omega(k)}$ terms. Thus, for $F = R$ and for every $F = R_i$,

$$\frac{1}{T}\sum_{t\in I} \chi(t)F(t)^2 = \rho\,\mathbb{E}F^2 + O\left(\frac{2^{\omega(k)}D^3}{T}\right). \tag{17}$$

Incompatible systems contribute zero to both sides, so they do not affect this bound.

Step 5: extracting and counting successful translates. Define

$$S_1 = \frac{1}{T}\sum_{t\in I} \chi(t)R(t)^2, \qquad S_2 = \frac{1}{T}\sum_{t\in I} \chi(t)R(t)^2 \sum_{i=1}^{r} \mathbf{1}_{(L_i(t),P)=1}.$$

Equations (13), (14), (15), and (17) imply

$$\begin{aligned} S_1 &= \rho(1+o(1))A^r + O(2^{\omega(k)}D^3/T), \\ S_2 &\ll \rho r A^r \left(e^{-u} + \frac{(\log D)^2}{z^2}\right) + r2^{\omega(k)}D^3/T. \end{aligned}$$

By (10), $A \geq 1$, and $3\kappa < \tau$, one has

$$\frac{r2^{\omega(k)}D^3}{\rho A^r T} \leq \exp\bigl(-(\tau - 3\kappa + o(1))x\bigr) = o(1).$$

The remaining relative errors tend to zero uniformly, since

$$re^{-u} \leq \delta x e^{-\kappa(\log x)^5} = o(1), \qquad r\frac{(\log D)^2}{z^2} = O(x^{3-2C}) = o(1).$$

Thus $S_1 - S_2 \geq \rho A^r/2$ for sufficiently large x.

Let $\mathcal{G} \subseteq I$ be the set of translates satisfying (9). For every $t \in I$,

$$\chi(t)R(t)^2\left(1 - \sum_i \mathbf{1}_{(L_i(t),P)=1}\right) \leq \mathbf{1}_{t\in\mathcal{G}}R(t)^2.$$

Also (16) gives $R(t)^2 \leq D^3$. Summing yields

$$|\mathcal{G}|D^3 \geq T(S_1 - S_2) \geq \tfrac{1}{2}\rho A^r T \geq \tfrac{1}{2}\rho T,$$

which proves the claimed count. Every auxiliary prime is at most y, so a positive form value greater than y with such a divisor is composite. The auxiliary divisors may depend on the successful translate; no common choice of them for all successful translates is required. □

3 A preliminary covering after forbidden primes are deleted

This section constructs a covering of progression indices, using only primes not dividing the modulus k. A single condition on the prime divisors of k controls both the loss from these forbidden primes and the number of indices left for Lemma 2.2.

Throughout this section, k is a positive integer, x is sufficiently large, and

$$L = \log x, \qquad L_2 = \log_2 x, \qquad L_3 = \log_3 x.$$

Define

$$V(k) = \prod_{p \mid k} (1 - p^{-3/4})^{-1}, \qquad \rho(k) = \frac{\varphi(k)}{k}.$$

Lemma 3.1 (Density-one forbidden-prime control)**.** *Let $x = \alpha \log K$, where $\alpha > 0$ is fixed, and let $K \to \infty$. All but $O(K/L_2)$ of the integers $K < k \le 2K$ satisfy $V(k) \le L_2$; in particular, this set has proportion $1 - o(1)$ in the dyadic interval. Moreover, for every sufficiently large x and every positive integer k satisfying $V(k) \le L_2$, one has*

$$\rho(k) \ge L_2^{-1}, \tag{18}$$

$$\sum_{\substack{p \mid k \\ p > L^4}} \frac{1}{p} \le \frac{L_3}{L} = o(1), \tag{19}$$

$$\sum_{\substack{p \mid k \\ p \le x}} \log p \le x^{3/4} L L_3 = o(x). \tag{20}$$

The three bounds are uniform in k and do not require a relation between x and k beyond $V(k) \le L_2$.

Proof. Write $v(p) = (1 - p^{-3/4})^{-1} - 1$ and extend v multiplicatively to squarefree integers, with $v(1) = 1$. Expanding the finite product gives $V(k) = \sum_{d \mid \mathrm{rad}(k)} v(d)$. Nonnegativity and $\lfloor 2K/d \rfloor \le 2K/d$ give

$$\frac{1}{K} \sum_{K < k \le 2K} V(k) \le 2 \sum_{d \text{ squarefree}} \frac{v(d)}{d} = 2 \prod_{p} \left(1 + \frac{v(p)}{p}\right) < \infty.$$

The Euler product converges because $v(p)/p = O(p^{-7/4})$; nonnegativity justifies its expansion and the preceding interchange of sums. Thus the displayed mean is bounded by an absolute constant C_V. Markov's inequality gives at most $C_V K/L_2$ exceptions, which proves the proportion statement since $L_2 \to \infty$.

For the remaining assertions, only $V(k) \le L_2$ is needed. Each factor in $\rho(k)^{-1}$ is at most the corresponding factor in $V(k)$, because $p^{-1} \le p^{-3/4}$. This proves (18). The inequality $t \le -\log(1-t)$ for $0 < t < 1$ gives

$$\sum_{p \mid k} p^{-3/4} \le \log V(k) \le \log L_2 = L_3.$$

For $p > L^4$, one has $p^{-1} \le L^{-1} p^{-3/4}$; for $p \le x$, one has $\log p \le x^{3/4} L\, p^{-3/4}$. Summing these two inequalities proves (19) and (20). Their right sides are respectively $o(1)$ and $o(x)$ as $x \to \infty$. □

Lemma 3.2 (Preliminary covering with forbidden primes)**.** *For every $\delta > 0$ there are constants $c = c(\delta) > 0$ and $x_0 = x_0(\delta)$ such that the following holds for every $x \ge x_0$ and every positive integer k with $V(k) \le L_2$. Set*

$$H = \left\lfloor c \rho(k) x \frac{L^2 L_3}{L_2^2} \right\rfloor.$$

One can choose one residue $a_p \pmod p$ for each prime $p \le x$ with $p \nmid k$ so that at most δx indices in $\{0, 1, \dots, H\}$ are outside the union of these classes.

Proof. Step 1: parameters and the initial zero residues. Temporarily fix any $0 < c \le 1$; its value will be chosen at the end. All estimates below are uniform in k with $V(k) \le L_2$. The numerical constants in the survivor bounds can be taken independent of c, although the threshold for x may depend on this fixed c. Put $w = L^4$ and set

$$u = \frac{5L_2}{L_3}, \qquad z_0 = H^{1/u}.$$

Since $L_2^{-1} \le \rho(k) \le 1$, the floor in H gives, for sufficiently large x,

$$\frac{c}{2}\frac{L^2 L_3}{L_2^3} \le \frac{H}{x} \le c\frac{L^2 L_3}{L_2^2}.$$

Consequently $H/x \to \infty$, $\log H = L + O_c(L_2)$, and $H < x^2$. Moreover,

$$\log z_0 = (1 + o(1))\frac{LL_3}{5L_2}, \qquad L_2 = o(\log z_0), \qquad \log z_0 = o(L).$$

Together with $H/x = o(L^4)$, these estimates imply

$$2H/x < w < z_0 < x/2.$$

Call a prime *allowed* if it does not divide k. Choose $a_p = 0$ at every allowed prime $p \le w$ and every allowed prime $z_0 < p \le x/2$. Let S_0 be the set of positive indices in $\{1, \dots, H\}$ surviving these choices.

Step 2: classifying and counting the survivors. For $n \in S_0$, factor $n = dm$ uniquely so that all prime factors of d divide k, and $(m, k) = 1$. If m has a prime factor $q > x/2$, then $m/q \le n/q < 2H/x < w$. Any prime factor of m/q would be allowed and at most w, contradicting survival. Thus $m = q$. If m has no such prime factor, all its prime factors are at most z_0, because an allowed prime in $(z_0, x/2]$ would already cover n. The latter case includes $m = 1$; an integer with no prime factors is regarded as z_0-smooth.

The number of indices of the first kind is bounded by

$$\sum_{\substack{d \le 2H/x \\ \mathrm{rad}(d) \mid k}} \pi(H/d) \ll \frac{H}{\log x} \sum_{\mathrm{rad}(d) \mid k} \frac{1}{d} = \frac{H}{\rho(k)L}.$$

Here $\pi(v)$ denotes the number of primes at most v. For every displayed d one has $H/d \ge x/2$, so the elementary bound $\pi(v) \ll v/\log v$ applies with $\log(H/d) \gg L$. The identity in the last step is the convergent geometric-series product $\sum_{\mathrm{rad}(d) \mid k} d^{-1} = \prod_{p \mid k}(1 - 1/p)^{-1} = \rho(k)^{-1}$.

Step 3: the smooth part, including forbidden prime factors. We use the elementary Rankin argument for smooth numbers; see also [6, Section 5]. For the second kind, put $\sigma = 1 - \log u/\log z_0$. Here $u > 1$ and $\log u/\log z_0 = O_c(L_2/L) = o(1)$, so $3/4 \le \sigma < 1$ for large x. Enlarge the count to all pairs of positive integers (d, m) with $dm \le H$, $\mathrm{rad}(d) \mid k$, and m z_0-smooth, dropping the condition $(m, k) = 1$. Applying $\mathbf{1}_{dm \le H} \le (H/(dm))^\sigma$ and summing the resulting geometric series gives the upper bound

$$H^\sigma \prod_{p \le z_0} (1 - p^{-\sigma})^{-1} \prod_{p \mid k} (1 - p^{-\sigma})^{-1}. \tag{21}$$

Primes dividing both k and $\prod_{p\le z_0} p$ can occur in both factors; this corresponds precisely to the enlargement of the pair count. The last product is at most $V(k) \le L_2$. Convexity of $t \mapsto u^t$ on $[0,1]$ gives

$$p^{1-\sigma} - 1 \le (u-1)\frac{\log p}{\log z_0} \quad (p \le z_0)$$

and Mertens' estimates therefore give

$$\prod_{p\le z_0} (1-p^{-\sigma})^{-1} \ll \log z_0\, e^{C_E u}$$

for an absolute constant $C_E > 0$. Indeed, the first powers in the logarithmic Euler expansion have difference at most

$$\frac{u-1}{\log z_0} \sum_{p\le z_0} \frac{\log p}{p} = O(u).$$

The higher powers are uniformly bounded, since

$$\sum_{p\le z_0} \sum_{\nu\ge 2} \frac{p^{-\nu\sigma}}{\nu} \le \frac{1}{1-2^{-3/4}} \sum_p p^{-3/2} < \infty.$$

This proves the product bound using $\prod_{p\le z_0}(1-1/p)^{-1} \ll \log z_0$. Now $\log H = u \log z_0$, so $H^\sigma = He^{-u\log u}$ exactly. Also

$$\log u = L_3 - \log L_3 + \log 5, \qquad u\log u = 5L_2 + o(L_2), \qquad u = o(L_2),$$

and $\log\log z_0 = L_2 + o_c(L_2)$. It follows that (21) is

$$\ll HL_2 \log z_0\, e^{-u\log u + C_E u} = HL^{-4+o(1)}L_2 = o\!\left(\frac{H}{\rho(k)L}\right).$$

For the final comparison, divide by $H/(\rho(k)L)$ and use $\rho(k) \le 1$: the ratio is at most $L_2 L^{-3+o(1)} = o(1)$. Combining the two types gives $|S_0| \le C_S H/(\rho(k)L)$ for an absolute constant C_S and all sufficiently large x (with the threshold allowed to depend on c).

Step 4: greedy covering and the final constant. At each unused allowed prime $w < p \le z_0$, choose a residue containing at least a $1/p$ fraction of the current survivors. This is possible by partitioning that set into the p residue classes; an empty survivor set needs no special choice. No prime in this range was used in Step 1. The resulting number of positive survivors is at most

$$|S_0| \prod_{\substack{w<p\le z_0\\ p\nmid k}} (1-1/p) \ll \frac{H}{\rho(k)L}\frac{\log w}{\log z_0} \ll \frac{HL_2^2}{\rho(k)L^2 L_3} \ll cx.$$

To justify the missing-prime correction explicitly, factor the product as

$$\prod_{\substack{w<p\le z_0\\ p\nmid k}} (1-1/p) = \prod_{w<p\le z_0} (1-1/p) \prod_{\substack{w<p\le z_0\\ p\mid k}} (1-1/p)^{-1}.$$

The first product is $\ll \log w/\log z_0$ by Mertens' theorem. The second lies between 1 and

$$\exp\left(2\sum_{\substack{p\mid k\\ p>w}} \frac{1}{p}\right) \le \exp(2L_3/L) = 1 + o(1),$$

by (19) and $-\log(1-1/p) \le 2/p$. Finally $\log w = 4L_2$ and $\log z_0 \sim LL_3/(5L_2)$, which proves the preceding survivor bound.

Thus there is an absolute constant $C_* > 0$ such that the number of positive survivors is at most $C_* cx$ for all sufficiently large x. Assign arbitrary residues to all remaining allowed primes $p \le x$; this cannot increase the survivor set. Including 0 among the indices adds at most one survivor, so the total is at most $C_* cx + 1$. Choose

$$c = \min\left(1, \frac{\delta}{2C_*}\right),$$

and then take x large enough for the preceding estimates and $x \ge 2/\delta$. The total is at most δx, as required. □

4 Direct transfer and multiplicity

We now combine the preliminary index cover with the reduced-residue short-translate lemma. The size of the preliminary modulus will ensure that distinct successful translates represent distinct reduced residue classes modulo k.

Proof of Theorem 1.1. Step 1: one density-one set for all fixed ϵ. Choose a fixed integer k_0 large enough that the iterated logarithms below are positive, and define

$$\mathcal{K} = \{k \ge k_0 : V(k) \le \tfrac{1}{2}\log_3 k\}.$$

This set has natural density one. Indeed, with v as in the proof of Lemma 3.1, the same nonnegative divisor expansion gives

$$\sum_{k\le X} V(k) \le X \sum_{d \text{ squarefree}} \frac{v(d)}{d} = C_{\mathrm{av}} X, \qquad C_{\mathrm{av}} = \prod_p \left(1 + \frac{v(p)}{p}\right) < \infty.$$

For sufficiently large X, every excluded integer k with $\sqrt{X} < k \le X$ satisfies $V(k) > \frac{1}{2}\log_3 \sqrt{X}$. Hence

$$\#\{k \le X : k \notin \mathcal{K}\} \le \sqrt{X} + \frac{2C_{\mathrm{av}} X}{\log_3 \sqrt{X}} = o(X).$$

In particular, the definition of $\mathcal{K}$ is independent of ϵ.

Step 2: parameters and the preliminary modulus. Fix $0 < \epsilon < 1$. In a dyadic interval $K < k \le 2K$, restrict to $k \in \mathcal{K}$ and take

$$x = \frac{\epsilon}{8}\log K, \quad \tau = \frac{8}{\epsilon} - 2, \quad T = \lfloor e^{\tau x} \rfloor, \quad D = e^{x/10}.$$

Write $L = \log x$, $L_2 = \log_2 x$, $L_3 = \log_3 x$, and $y = \exp(x/L^5)$. Uniformly for $K < k \le 2K$,

$$L_2 = \log_2\big((\epsilon/8)\log K\big) \sim \log_3 k.$$

Thus every $k \in \mathcal{K}$ in this interval satisfies $V(k) \le L_2$ once K is sufficiently large in terms of ϵ. All subsequent asymptotic estimates are uniform in these k, with ϵ fixed.

Choose the parameters of Lemma 2.2 as

$$A_0 = 2, \qquad C = 7, \qquad \kappa = 1/10,$$

and fix $C_0 > 8/\epsilon + 2$. The inequalities $C > \max(A_0, 6)$ and $0 < \kappa < \tau/3$ hold because $7 > 6$ and $\tau > 6$. Let $\delta > 0$ be supplied by that lemma for these fixed parameters. Define

$$M = \prod_{\substack{p \le x \\ p \nmid k}} p.$$

Then $(M, k) = 1$. Writing $\vartheta(x) = \sum_{p \le x} \log p$, the prime number theorem and (20) give

$$\log M = \vartheta(x) - \sum_{\substack{p \mid k \\ p \le x}} \log p = (1 + o(1))x.$$

Consequently $M > \max(x, y)$ for large K. Moreover,

$$\log(kM) \le \log(2K) + (1 + o(1))x = (8/\epsilon + 1 + o(1))x < C_0 x,$$

which verifies the quantitative modulus hypothesis of Lemma 2.2. Since $\log(T + 1) = \tau x + o(1)$, one also has

$$\log\big(M(T + 1)\big) = (1 + \tau + o(1))x = \left(1 - \frac{\epsilon}{8} + o(1)\right) \log K < \log K,$$

so $M(T + 1) < K < k$ for all sufficiently large K.

Step 3: transferring the index cover to progression terms. Apply Lemma 3.2 with this value of δ, and denote its constant by $c_{\mathrm{cov}} > 0$. Thus

$$H = \left\lfloor c_{\mathrm{cov}} \rho(k) x \frac{L^2 L_3}{L_2^2} \right\rfloor.$$

Let $a_p \pmod p$ be the chosen covering residues, and let $J \subseteq \{0, \dots, H\}$ be the uncovered indices. Then $|J| \le \delta x$, and the parameter estimates in the proof of Lemma 3.2 give $H < x^2 = x^{A_0}$. Hence J has the span required by Lemma 2.2. The Chinese remainder theorem gives a unique b with $0 \le b < M$ and

$$b \equiv -k a_p \pmod p \qquad (p \le x,\ p \nmid k).$$

Apply Lemma 2.2 on the interval $I = \{1, \dots, T\}$, and let $\mathcal{T} \subseteq I$ be its set of successful translates. The preceding estimates verify all its hypotheses, including the coprimality and modulus-size conditions and the cardinality and span of J. It gives

$$|\mathcal{T}| \ge \frac{\rho(k) T}{2D^3}.$$

For $j \in \{0, \dots, H\} \setminus J$, some allowed prime $p \le x$ has $j \equiv a_p \pmod p$. As $p \mid M$, for every $t \in I$ we obtain

$$b + Mt + kj \equiv k(j - a_p) \equiv 0 \pmod p.$$

The value is at least $M > x$, so this divisor is proper. For each $t \in \mathcal{T}$, Lemma 2.2 supplies an auxiliary prime divisor at most y for every form indexed by $j \in J$. These values are also at least $M > y$, so their divisors are proper. This argument includes the index $j = 0$, whether it belongs to J or not.

For $t \in \mathcal{T}$, put $a = b + Mt$. The inequalities

$$1 \le M \le a \le b + MT < M(T + 1) < k$$

and the coprimality conclusion of Lemma 2.2 give

$$1 \le a < k, \qquad (a,k) = 1, \qquad a, a+k, \dots, a+kH \text{ are all composite.}$$

Because $1 \le a < k$, every positive integer in this residue class has the form $a + kj$ with $j \ge 0$. The displayed terms are therefore all its positive members up to $a + kH$. Since they are composite and $(a,k) = 1$, the least prime in the class satisfies

$$p(k,a) > a + kH > kH.$$

Step 4: multiplicity and comparison of scales. Distinct $t \in \mathcal{T}$ give distinct integers $a = b + Mt$, since $M > 0$; as all these integers lie in $[1,k)$, their residue classes modulo k are distinct as well. For sufficiently large K, $T \ge (2/3)e^{\tau x}$. Using $\rho(k) \ge L_2^{-1}$ yields

$$|\mathcal{T}| \ge \frac{\rho(k)T}{2D^3} \ge \frac{1}{3L_2} K^{\,1-23\epsilon/80} \ge k^{1-\epsilon}.$$

Here $(\tau - 3/10)x = (1 - 23\epsilon/80)\log K$, and the last inequality follows from $k \le 2K$ and

$$\frac{K^{1-23\epsilon/80}}{3L_2(2K)^{1-\epsilon}} = \frac{K^{57\epsilon/80}}{3\,2^{1-\epsilon}L_2} \longrightarrow \infty.$$

The floor in H is negligible uniformly, since its defining expression is at least $c_{\mathrm{cov}} x L^2 L_3 / L_2^3 \to \infty$. Thus $H \sim c_{\mathrm{cov}} \rho(k) x L^2 L_3 / L_2^2$, and

$$\frac{x}{\log k} \longrightarrow \frac{\epsilon}{8}, \quad L \sim \log_2 k, \quad L_2 \sim \log_3 k, \quad L_3 \sim \log_4 k.$$

Recalling the definition of F, we obtain the uniform limit

$$\frac{kH}{\varphi(k)F(k)} \longrightarrow \frac{\epsilon c_{\mathrm{cov}}}{8}.$$

Taking $c_\epsilon = \epsilon c_{\mathrm{cov}}/16 > 0$, every successful class consequently satisfies $p(k,a) > kH \ge c_\epsilon \varphi(k)F(k)$ for all sufficiently large K. Together with the count of distinct classes, this proves the first assertion for every sufficiently large $k \in \mathcal{K}$, by taking the dyadic intervals $K = 2^m$. The constants and thresholds may depend on ϵ, but the set $\mathcal{K}$ does not. Finally fix $\epsilon = 1/2$ and select one of the resulting classes. Then $P(k) \ge c_{1/2}\varphi(k)F(k)$ for every sufficiently large $k \in \mathcal{K}$, proving the final assertion with an absolute positive constant. □

Use of AI: The author wrote a draft (without using AI) explaining how to improve $P(k)$ using the techniques from [5, 16]. The author then fed this draft to multiple AI tools (Claude, ChatGPT and two open-weight models with open-source harnesses) and prompted them to work out the details of the proofs and identify missing gaps. The author fixed those missing gaps and iteratively verified and expanded the proofs using AI tools. The author is solely responsible for the correctness of the proofs.

References

[1] S. Chowla, *On the least prime in an arithmetical progression*, J. Indian Math. Soc. **1** (1934), 1–3.

[2] P. Erdős, *On the difference of consecutive primes*, Quart. J. Math. Oxford Ser. **6** (1935), 124–128.

[3] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao, *Long gaps between primes*, J. Amer. Math. Soc. **31** (2018), 65–105. https://doi.org/10.1090/jams/876.

[4] A. Fiori, *The least prime in arithmetic an progression*, preprint, 2024. https://arxiv.org/abs/2404.02329.

[5] OpenAI, *Improved long gaps between primes*, preprint, 2026. Online manuscript.

[6] K. Ford, *Sieve methods lecture notes*, Spring 2023, Theorem 4.4 and Section 5. https://ford126.web.illinois.edu/sieve2023.pdf.

[7] A. Granville and C. Pomerance, *On the least prime in certain arithmetic progressions*, J. London Math. Soc. (2) **41** (1990), 193–200.

[8] D. Hensley and I. Richards, *Primes in intervals*, Acta Arith. **25** (1974), 375–391.

[9] U. V. Linnik, *On the least prime in an arithmetic progression. I. The basic theorem*, Rec. Math. (Mat. Sbornik) N.S. **15(57)** (1944), 139–178.

[10] J. Li, K. Pratt, and G. Shakan, *A lower bound for the least prime in an arithmetic progression*, Quart. J. Math. **68** (2017), 729–758. https://doi.org/10.1093/qmath/hax001.

[11] C. Pomerance, *A note on the least prime in an arithmetic progression*, J. Number Theory **12** (1980), 218–223.

[12] K. Prachar, *Über die kleinste Primzahl einer arithmetischen Reihe*, J. Reine Angew. Math. **206** (1961), 3–4.

[13] R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.

[14] A. Schinzel, *Remark on the paper of K. Prachar "Über die kleinste Primzahl einer arithmetischen Reihe"*, J. Reine Angew. Math. **210** (1962), 121–122.

[15] Y.-C. Sun and H. Pan, *On the gaps between consecutive primes*, preliminary preprint, 2018. https://arxiv.org/abs/1802.02470.

[16] GPT 5.6 Sol, *A tilted residue-class construction for long prime-free intervals*, preprint, 25 August 2026. https://github.com/DottedCalculator/ai-math/blob/main/Erdos_4_GPT_5.6_Sol.pdf.

[17] S. S. Wagstaff, Jr., *The least prime in an arithmetic progression with prime difference*, J. Reine Angew. Math. **301** (1978), 114–115.

[18] S. S. Wagstaff, Jr., *Greatest of the least primes in arithmetic progressions having a given modulus*, Math. Comp. **33** (1979), 1073–1080. https://doi.org/10.2307/2006082.

[19] T. Xylouris, *On the least prime in an arithmetic progression and estimates for the zeros of Dirichlet L-functions*, Acta Arith. **150** (2011), 65–91. https://doi.org/10.4064/aa150-1-4.

[20] T. Xylouris, *Linniks Konstante ist kleiner als* 5, Chebyshevskii Sb. **19** (2018), no. 3, 80–94. https://doi.org/10.22405/2226-8383-2018-19-3-80-94.