BALANCED SEMIPRIMES IN SHORT INTERVALS UNDER THE RIEMANN HYPOTHESIS

ADAM HENRIKSSON

Abstract. Assuming the Riemann hypothesis, we show that for every real $x \geq 9.6 \cdot 10^{37}$ the open interval of length $384\sqrt{x}(\log x)^2$ above x contains at least $(14/3)\sqrt{x}$ distinct products of two primes of size comparable to $\sqrt{x}$. Their contribution to the sum of the least prime factor divided by the integer is at least $14/9$. For every fixed $C > 0$, we also prove that the composite sum in an interval of length $C\sqrt{x}(\log x)^2$ is bounded above unconditionally and bounded away from zero under the Riemann hypothesis. We present a computer-assisted proof, verified in Lean 4, that the mean-square estimate of Chirre and Hagen implies the explicit count and weight bounds. The results have not been independently reviewed.

hexagon:2611.00808v1

1. Introduction

Let $P^-(n)$ be the least prime factor of an integer $n \geq 2$. Erdős and Graham [6, p. 92] ask whether the sum of $P^-(n)/n$ in an interval of length $C\sqrt{x}(\log x)^2$ above x is bounded away from zero for some $C > 0$ and all sufficiently large x. The interval sum in their text runs over all integers, although the preceding sum excludes primes. Tao [9] notes that this distinction matters. A prime contributes 1 by itself, and under the Riemann hypothesis Cramér's prime-gap bound supplies a prime in every sufficiently large interval of this length [4, 5]. This implication for the all-integer sum is already recorded in [7].

For the composite-only interpretation, Tao [9] points to products pq in which both primes have size $x^{1/2}(\log x)^{O(1)}$, and Zhang [11, §1.1] records this observation. We count products in a narrower range, with both primes within fixed multiples of $\sqrt{x}$. If $q < p$ then $P^-(pq)/(pq) = 1/p$, so a lower bound for their number gives a lower bound for the composite sum. Throughout, RH denotes the classical Riemann hypothesis. For real $x > 1$ and $C > 0$, write

$$H_C(x) = C\sqrt{x}(\log x)^2, \qquad \mu_C(x) = \sum_{\substack{x \leq n \leq x + H_C(x) \\ n \text{ composite}}} \frac{P^-(n)}{n}.$$

All logarithms are natural.

Theorem 1.1. *Assume RH. For every real $x \geq 9.6 \cdot 10^{37}$, at least $(14/3)\sqrt{x}$ distinct integers in $(x, x + H_{384}(x))$ have the form pq, where p, q are primes and*

$$\sqrt{x}/2 < q < \sqrt{x} < p < 3\sqrt{x}.$$

These integers alone contribute at least $14/9$ to $\mu_{384}(x)$.

The count gives an explicit lower bound for the composite sum at $C = 384$. For a general fixed $C > 0$ the bounds take the following form.

Theorem 1.2. *There is an absolute constant K such that, for each fixed $C > 0$, we have $\mu_C(x) \leq K(1 + C)$ for all sufficiently large x depending on C. Under RH, for each fixed $C > 0$ there are constants $c_C > 0$ and x_C such that $\mu_C(x) \geq c_C$ for every real $x \geq x_C$.*

Zeraoulia [10, Proposition 4] states a lower bound for intervals of length $C\sqrt{x}(\log x)^B$, with $B > 2$, under a lower bound for the number of primes in each factor interval of length $(\log Y)^B$. At exponent 2, Zhang [11, Hypothesis 1.5 and Theorem 1.7] obtains two-sided bounds from uniform upper and lower estimates for prime counts in short intervals. Our proof keeps x fixed and varies the two intervals in which the factors are sought. An estimate for the mean square of the prime-counting error then suffices.

Date: 21 November 2026.

2. The explicit semiprime count

Write $\theta(y) = \sum_{p \le y} \log p$, with p running over primes, and let $\pi(y)$ be the number of primes up to y. For $Y \ge 1$ and $\delta > 0$ put

$$J_\theta(Y, \delta) = \int_1^Y \big(\theta((1+\delta)v) - \theta(v) - \delta v\big)^2 \, dv.$$

Chirre and Hagen [2, Theorem 3] prove under RH that

$$J_\theta(Y, \delta) \le 2.5571\, \delta Y^2 \log^2(2.0001/\delta) \qquad (Y \ge 10^{13}, \quad 0 < \delta \le 10^{-13}). \tag{1}$$

They apply this estimate to sums of two primes. For products, the two intervals start at t and x/t. The point of the proof is to find many t for which both intervals contain enough primes.

Proof of Theorem 1.1. Put $N = \sqrt{x}$, $L = \log N$ and $\delta = 512L^2/N$. We first check the range of (1). Set $B = 9797 \cdot 10^{15}$. Since $B^2 < 9.6 \cdot 10^{37}$, we have $N \ge B$. The rational inequalities

$$B^{15} < (1359/500)^{656}, \qquad 1359/500 < \sum_{j=0}^{6} 1/j! < e$$

give $\log B < 656/15$. The function $(\log N)^2/N$ decreases for $N > e^2$, so

$$0 < \delta \le \frac{512(\log B)^2}{B} < \frac{512(656/15)^2}{B} < 10^{-13}.$$

Both $2N$ and $2N/3$ exceed 10^{13}. Also $2.0001 < 512L^2 = \delta N$, so $0 < \log(2.0001/\delta) \le L$. The only two estimates we need are therefore

$$J_\theta(Y, \delta) \le 3\delta Y^2 L^2, \qquad Y = 2N,\ 2N/3. \tag{2}$$

Let $I = [3N/2, 2N]$ and $J = [N/2, 2N/3]$. The map $t \mapsto x/t$ takes I onto J. Put $u = 1 + \delta$. In each interval call v exceptional if $\theta(uv) - \theta(v) < \delta v/2$, and denote the two exceptional sets by B_I and B_J. The absolute error $|\theta(uv) - \theta(v) - \delta v|$ is at least $3\delta N/4$ on B_I and at least $\delta N/4$ on B_J. Writing $|E|$ for Lebesgue measure, we obtain from (2)

$$|B_I| \le \frac{3\delta(2N)^2 L^2}{(3\delta N/4)^2} = \frac{N}{24}, \qquad |B_J| \le \frac{3\delta(2N/3)^2 L^2}{(\delta N/4)^2} = \frac{N}{24}.$$

For $v \in J$ the inverse map has derivative of absolute value $x/v^2 \le 4$. Thus at most $4|B_J|$ of I, in measure, maps into B_J. The set G of $t \in I$ for which neither t nor x/t is exceptional satisfies

$$|G| \ge N/2 - |B_I| - 4|B_J| \ge 7N/24. \tag{3}$$

Let $A(v) = \pi(uv) - \pi(v)$. For $t \in G$ and $v = t$ or $v = x/t$, the primes counted by $A(v)$ are below $3N$. Hence $A(v)\log(3N) \ge \theta(uv) - \theta(v) \ge \delta v/2$. Using $t(x/t) = N^2$ and $\log(3N) \le 2L$ gives

$$\int_I A(t)A(x/t)\, dt \ \ge\ |G| \frac{\delta^2 N^2}{4\log^2(3N)} \ \ge\ \frac{7N}{24}\,\frac{\delta^2 N^2}{16L^2}. \tag{4}$$

For a fixed pair (p, q), the conditions $t < p \le ut$ and $x/t < q \le ux/t$ hold precisely on

$$I \cap [p/u, p) \cap (x/q, ux/q].$$

A pair counted anywhere in I has $p \le 2uN$. Its contribution to the integral is therefore at most $p - p/u = \delta p/u \le 2\delta N$. Moreover,

$$N/2 < q \le 2uN/3 < N < 3N/2 < p \le 2uN < 3N.$$

Thus the larger factor determines the order of the pair, and unique factorization gives a different integer for each pair. Dividing (4) by $2\delta N$ bounds the number of these integers below by

$$\frac{7N}{24}\,\frac{\delta^2 N^2}{16L^2}\,\frac{1}{2\delta N} = \frac{14}{3}N.$$

Since $0 < \delta < 1$, each product satisfies

$$x < pq \le (1+\delta)^2 x < x + 3\delta x = x + H_{384}(x).$$

Finally $P^-(pq)/(pq) = 1/p \ge 1/(3N)$. Summing over the products gives 14/9. □

The supplement gives another choice of explicit constants from the mean-square estimate of Cully-Hugill and Dudek [3, Theorem 1]. For $x \ge 10^{42}$ it gives at least $1024\sqrt{x}$ balanced products in an interval of length $98304\sqrt{x}(\log x)^2$ above x.

Formal verification. Lean verifies the implication from (1) to Theorem 1.1, with (1) supplied as a hypothesis. Its derivation from RH is the published theorem of Chirre and Hagen. A second formal proof starts from their estimate for ψ [2, Theorem 2] at the two values of Y in (2). The passage from prime powers to primes in that proof adapts Hubbard's LeanLab code [8]. The proof of Theorem 1.2 in Sections 3 and 4 has not been formalized in Lean.

3. The lower bound for fixed C

For small C we allow the two primes to have different sizes. We use the following unconditional estimate for the smaller prime.

Lemma 3.1. *Put $A_h(s) = \pi(s+h) - \pi(s)$. There is an absolute S_0 such that for all $S \ge S_0$ and all real h with $2\log S \le h \le S/4$,*

$$\int_S^{2S} A_h(s)^2\, ds \le 132 S h^2/(\log S)^2. \tag{5}$$

The set of $s \in [S, 2S]$ for which $A_h(s) \ge h/(4\log S)$ has measure at least bS, where $b = 1/2500$.

Proof. Let $\psi(y) = \sum_{p^a \le y} \log p$, summed over primes p and integers $a \ge 1$. We use Gallagher's moment estimate in the form proved by Bazzanella, Languasco and Zaccagnini [1, Lemma 4] using Klimov's sieve bound. With $k = 2$, $\varepsilon = 1$ and $P_2(z) = 2z + 8z^2$, it gives

$$\sum_{m \le X} (\psi(m+j) - \psi(m))^2 \le X\big((\log X)^2 + 2j\log X + 8j^2\big)$$

for all $X \ge X_0$ and every integer $1 \le j \le X$. Here we use Lemma 4 with a single threshold $X_0 = X_0(k, \varepsilon)$ for all $1 \le j \le X$. The authors state the range of j and leave the dependence of the sufficiently large X implicit. Theorem 1.2 uses only the existence of a uniform second-moment constant and a positive proportion of good starts. Take $X = \lceil 2S \rceil$ and $j = \lceil h \rceil + 1$, so $\log X \le j \le 2h$ and $j \le X$. For $m = \lfloor s \rfloor$, each prime counted in $A_h(s)$ has weight at least $\log S$ in $\psi(m+j) - \psi(m)$. The set of s with $\lfloor s \rfloor = m$ has length at most one. Summing over m gives (5) since $11Xj^2 \le 132Sh^2$.

Each prime in $(5S/4, 2S]$ contributes h to the first moment. The prime number theorem thus gives

$$\int_S^{2S} A_h(s)\, ds \ge h\big(\pi(2S) - \pi(5S/4)\big) \ge Sh/(2\log S).$$

Let $E = \{s \in [S, 2S] : A_h(s) \ge h/(4\log S)\}$. Its complement contributes at most $Sh/(4\log S)$, so $\int_E A_h(s)\, ds \ge Sh/(4\log S)$. Cauchy–Schwarz and (5) now give $|E| \ge S/(16 \cdot 132) = S/2112 > bS$. □

Proof of the lower bound in Theorem 1.2. Given $C > 0$, fix $r \ge \max(2, 192/(bC))$. Write $N = \sqrt{x}$, $L = \log N$ and set

$$S = N/(2r), \qquad T = rN, \qquad \eta = CL^2/N, \qquad h = \eta S.$$

Since r is fixed and $h = CL^2/(2r)$, all sufficiently large N satisfy $2\log S \le h \le S/4$. For $S \le s \le 2S$, the second interval starts at $v = x/s \in [T, 2T]$, since $x = 2ST$. Apply (1) at $Y = 2T$ and $\delta = \eta$. We have $0 < \eta \le 10^{-13}$, $Y \ge 10^{13}$ and $0 < \log(2.0001/\eta) \le L$. The set of $v \in [T, 2T]$ where $\theta((1+\eta)v) - \theta(v) < \eta v/2$ has measure at most

$$3\eta(2T)^2 L^2/(\eta T/2)^2 = 48L^2/\eta.$$

Under $s = x/v$ this set has preimage of measure at most

$$\frac{x}{T^2}\frac{48L^2}{\eta} = \frac{96S}{Cr} \le \frac{bS}{2}.$$

Lemma 3.1 gives a set of measure at least bS on which $A_h(s) \ge h/(4\log S)$. After removing this preimage, a set of measure at least $bS/2$ remains.

For each such s there are at least $h/(4L)$ primes $q \in (s, s+h]$ and at least $\eta T/(4L)$ primes $p \in (v, (1+\eta)v]$, since $\log S \le L$ and $\log(3T) \le 2L$. A fixed pair (q,p) can occur only for $q - h \le s < q$, an interval of length h. Moreover $q \le 2S + h \le 9N/(8r) < N$ and $N < p \le (1+\eta)2T < 3T$. Integrating the product of the two counts over the good set and dividing by h gives at least

$$bS\eta T/(32L^2) = bCN/64$$

distinct products. Since $h \le \eta s$, each lies in $(x, (1+\eta)^2 x]$. For $\eta \le 1$, the excess is at most $3\eta x = 3CNL^2 \le H_C(x)$. Each has weight $1/p \ge 1/(3T)$. Hence $\mu_C(x) \ge bC/(192r)$. □

4. The unconditional upper bound

Proof of the upper bound in Theorem 1.2. With $N = \sqrt{x}$ and $L = \log N$, put $H = H_C(x)$, $d = H/x$ and $y = N/(CL)$. A prime $p \le y$ divides at most $H/p + 1$ integers in $[x, x+H]$. By the prime number theorem $\pi(y) \ll y/\log y$, and $\log y \sim L$ for fixed C. The total weight from least prime factors at most y is therefore at most

$$\sum_{p \le y} \frac{p}{x}\left(\frac{H}{p} + 1\right) \ll d\frac{y}{\log y} + \frac{y^2}{x \log y} \ll 1$$

for large x depending on C. For sufficiently large x, we have $y^3 > x + H$. Every remaining composite thus has exactly two prime factors, counted with multiplicity. Write it as pq with $y < p \le q$. Group the smaller primes in intervals $(S, 2S]$, where $S = 2^j y$, $j \ge 0$, and put $T = x/S$. Only values $S \le \sqrt{x+H} \le 2N$ can contribute.

For $t \in [S/2, 2S]$ write $v = x/((1+d)t)$. Let $A(t)$ count primes in $(t, (1+d)t]$ and $B(v)$ count primes in $(v, (1+d)^2 v]$. If $S < p \le 2S$ and $x \le pq \le (1+d)x$, this pair contributes to $A(t)B(v)$ for every $p/(1+d) < t < p$. Indeed $q \ge x/p > v$ and $q \le (1+d)x/p < (1+d)^2 v$. For $d \le 1$ this interval lies in $[S/2, 2S]$ and its length is $dp/(1+d) \ge dS/2$. This also covers $p = q$ and products at either endpoint.

On this range the first prime interval has length at most $2dS$. The variable v lies in $[T/4, 2T]$, its prime interval has length at most $6dT$, and $|dt/dv| \le 16S/T$. Since $S \ge N/(CL)$ and $S \le 2N$, we have $dS \ge 4L$ and $dT \ge 2CL^2$. Split the two ranges into intervals $[U, 2U]$ starting at $S/2, S$ and $T/4, T/2, T$, respectively. The bounds on S imply $L/2 \le \log U \le 2L$ uniformly for large x depending on C. We may take $d \le 1/96$ and $3CL \ge 1$. Then $2dS$ lies between $4L$ and $S/8$, and $6dT$ between $4L$ and $T/16$. On every dyadic interval the respective fixed length thus lies in $[2\log U, U/4]$. Lemma 3.1 and the inverse derivative bound give

$$\int_{S/2}^{2S} A(t)^2\,dt \ll S(dS)^2/L^2, \qquad \int_{S/2}^{2S} B\Big(\frac{x}{(1+d)t}\Big)^2\,dt \ll S(dT)^2/L^2.$$

If M_S is the number of pairs in this band, the interval of length at least $dS/2$ found above gives

$$\frac{dS}{2}\,M_S \le \int_{S/2}^{2S} A(t)B\Big(\frac{x}{(1+d)t}\Big)\,dt \ll \frac{S(dS)(dT)}{L^2}$$

by Cauchy–Schwarz. It follows that $M_S \ll dST/L^2 = 4CN$. Each product has weight $1/q = p/(pq) \le 2S/x$. The values of S form a geometric progression with sum at most $4N$, so these bands contribute $O(C)$ in all. Together with the $O(1)$ from $p \le y$, this proves the upper bound. □

Appendix A. Supplementary proofs

A.1. **A disjoint counting proof with the constants of Theorem 1.1.** Let N, I, δ and A be as in the proof of Theorem 1.1, and put $w = 2\delta N$ and $F(t) = A(t)A(x/t)$. Let M be the number of distinct products in the interval and prime ranges of that theorem. For each $0 \le a < w$ consider the starts $t \in I \cap (a + w\mathbb{Z})$. The intervals $(t, (1+\delta)t]$ at these starts are disjoint. Indeed, if $t' > t$ is another start, then $t' \ge t + w \ge (1+\delta)t$. The open left endpoint also takes care of equality. At a single start the two prime ranges are disjoint, so unique factorization counts $F(t)$ different products. Across different starts the larger primes are distinct. Thus $\sum_t F(t) \le M$ at every shift, including shifts on interval boundaries.

Integrating in a and translating each lattice term gives

$$wM \;\ge\; \int_0^w \sum_{k\in\mathbb{Z}} \mathbf{1}_I(a+kw)F(a+kw)\,da \;=\; \int_I F(t)\,dt \;\ge\; \frac{7N}{24}\,\frac{\delta^2N^2}{16(\log N)^2}.$$

Here the sum is understood as restricted to $a + kw \in I$, so only finitely many terms arise. Dividing by w gives $M \ge (14/3)N$. Each product has weight at least $1/(3N)$, so their total weight is at least $14/9$. Both endpoints are strict: $x < pq \le (1+\delta)^2 x < (1+3\delta)x$ because $0 < \delta < 1$.

At the cutoff of Theorem 1.1, the restriction $\delta \le 10^{-13}$ is nearly binding. On the decreasing branch $N > e^2$, the unique solution of $512(\log N)^2/N = 10^{-13}$ lies between 9790103351710472581 and 9790103351710472582. Squaring these bounds shows that $9.6 \cdot 10^{37}$ exceeds the corresponding threshold in x by about 0.16055%. The verifier bounds the logarithms with the series for $2\operatorname{arctanh} z$ and a geometric bound for its tail.

A.2. **A fixed-length argument.** Let $N = \sqrt{x}$, $I = [3N/2, 2N]$ and $J = [N/2, 2N/3]$. The map $t \mapsto x/t$ takes I onto J, and the inverse derivative has absolute value at most 4. Write $P^-(n)$ for the smallest prime factor of n. We prove the existence of admissible constants using fixed-length factor intervals. For some absolute D, Proposition 1 of Grenié, Molteni and Perelli [12], specialized to the number field $\mathbb{Q}$, gives under RH

$$\int_Y^{2Y} |\psi(v+h) - \psi(v) - h|^2\,dv \le DY(h + \log^2 Y)\log^2 Y$$

for all sufficiently large Y and $2 \le h \le Y$. Here $\psi(y) = \sum_{p^a \le y} \log p$.

Take $h = K\log^2 N$ with fixed $K \ge \max(1, 192D)$. On each of $[N, 2N]$ and $[N/2, N]$, the starts with a ψ-increment below $h/2$ have measure at most $8DY/K$, with $Y = N, N/2$ respectively. On I, the union of these exceptions and the reciprocals of the second set has measure at most $24DN/K \le N/8$. Let G be the set of starts in I giving two large increments. Then $|G| \ge N/4$.

For $v \in [N/2, 2N]$, the interval $(v, v+h]$ has at most one integer a-th power for each $a \ge 2$, once N is sufficiently large. Indeed consecutive a-th powers in the interval would be separated by at least $(N/2)^{(a-1)/a} \ge \sqrt{N/2} > h$. The total higher-prime-power weight in such an interval is therefore at most

$$\log(3N) \sum_{2 \le a \le \log(3N)/\log 2} \frac{1}{a} = O(\log N \log\log N) = o(h).$$

Each good interval then contains at least $h/(4\log(3N))$ primes. Choose a shift $a \in [0, h)$ for which the grid $a + h\mathbb{Z}$ meets G at least $N/(4h)$ times. Such a shift exists because

$$\int_0^h \sum_{k\in\mathbb{Z}} \mathbf{1}_G(a+kh)\,da = |G|.$$

The intervals $(t, t+h]$ for these grid starts are disjoint. At each start there are at least $h^2/(16\log^2(3N))$ pairs. For large N we have $x/t + h < N < t$ and $t + h < 3N$. Different starts give different larger prime factors, and each product has a unique factorization. There are at least $Nh/(64\log^2(3N))$ distinct products. Each belongs to $(x, x + 4Nh]$, since

$(t+h)(x/t+h) = x+h(t+x/t)+h^2$, $t+x/t \le 8N/3$, and $h \le N$ for large N. Each contributes at least $1/(3N)$. As $\log(3N) \le 2\log N$, the total weight is at least $K/768$, uniformly in all sufficiently large x, and $4Nh = K\sqrt{x}(\log x)^2$.

A.3. **A second explicit choice of constants.** Under RH, Theorem 1 of Cully-Hugill and Dudek [3] gives

$$\int_1^Y |\theta((1+\delta)v) - \theta(v) - \delta v|^2 \, dv < 202\delta Y^2 \log^2 Y$$

for $Y \ge 10^8$ and $0 < \delta \le 10^{-8}$. Fix $x \ge 10^{42}$. In the proof of Theorem 1.1, use $\delta = 131072L^2/N$ with $L = \log N$. The function L^2/N decreases for $N > e^2$ and $131072 \cdot 63^2/10^{21} < 10^{-8}$. For $Y = 2N, 2N/3$ one has $\log Y \le 2L$. Thus the coefficient in the mean-square bound $3\delta Y^2 L^2$ can be replaced by 808. The same applications of Chebyshev give $|B_I|, |B_J| \le (64 \cdot 808/9)L^2/\delta$. At least

$$N\left(\frac{1}{2} - \frac{320 \cdot 808}{9 \cdot 131072}\right) > N/4$$

units of t give two good starts. The incidence argument then gives at least $\delta N^2/(128L^2) = 1024N$ distinct products, with weight at least $1024/3$, in $(x, x + 98304\sqrt{x}(\log x)^2]$.

Use of AI tools

The research, the programs and a draft of this paper were developed with the assistance of AI models. The main AI contributions to the proof came from GPT-6.2 Astra. The author is responsible for the content.

References

[1] D. Bazzanella, A. Languasco and A. Zaccagnini, Prime numbers in logarithmic intervals, *Trans. Amer. Math. Soc.* **362** (2010), 2667–2684.

[2] A. Chirre and M. V. Hagen, On Goldbach numbers in short intervals, *Res. Number Theory* **12** (2026), article 82.

[3] M. Cully-Hugill and A. W. Dudek, An explicit mean-value estimate for the prime number theorem in intervals, *J. Aust. Math. Soc.* **117** (2024), 1–15.

[4] H. Cramér, Some theorems concerning prime numbers, *Ark. Mat. Astron. Fys.* **15** (1920), no. 5, 1–33.

[5] A. W. Dudek, On the Riemann hypothesis and the difference between primes, *Int. J. Number Theory* **11** (2015), 771–778.

[6] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Monographies de L'Enseignement Mathématique, vol. 28, Geneva, 1980.

[7] `gpt-erdos` project, Problem 462, `candidate_solution.md`, revision `7f63383`.

[8] S. Hubbard, LeanLab source for the comparison of θ and ψ, `ThetaPsi.lean`, revision `74c2d1a`, 2026.

[9] T. Tao, Comment on Erdős Problem 462, 28 September 2025.

[10] R. Zeraoulia, Toward Erdős–Graham Problem 462 (composite-only): exact rough-number and semiprime decompositions, averaged progress, and quantitative necessary conditions, preprint, 3 February 2026.

[11] X. Zhang, On the sum of least prime factors in short intervals, arXiv:2608.24930v2, 2026.

[12] L. Grenié, G. Molteni and A. Perelli, Primes and prime ideals in short intervals, *Mathematika* **63** (2017), 364–371.