Journal of Number Theory 185 (2018) 93–120

Contents lists available at ScienceDirect

Journal of Number Theory

www.elsevier.com/locate/jnt

Primitive root biases for prime pairs I: Existence and non-totality of biases ☆

Stephan Ramon Garcia [a,*], Florian Luca [b,c,d], Timothy Schaaff [a]

[a] *Department of Mathematics, Pomona College, 610 N. College Ave., Claremont, CA 91711, United States*
[b] *School of Mathematics, University of the Witwatersrand, Private Bag 3, Wits 2050, Johannesburg, South Africa*
[c] *Max Planck Institute for Mathematics, Vivatgasse 7, 53111 Bonn, Germany*
[d] *Department of Mathematics, Faculty of Sciences, University of Ostrava, 30 dubna 22, 701 03 Ostrava 1, Czech Republic*

A R T I C L E   I N F O

Article history:
Received 16 August 2017
Received in revised form 29 October 2017
Accepted 31 October 2017
Available online 17 November 2017
Communicated by S.J. Miller

MSC:
11A07
11A41
11N36
11N37

Keywords:
Prime
Twin prime
Primitive root

A B S T R A C T

We study the difference between the number of primitive roots modulo p and modulo $p+k$ for prime pairs $p, p+k$. Assuming the Bateman–Horn conjecture, we prove the existence of strong sign biases for such pairs. More importantly, we prove that for a small positive proportion of prime pairs $p, p+k$, the dominant inequality is reversed.

☆ SRG supported by NSF grant DMS-1265973 and the Budapest Semesters in Mathematics (BSM) Director's Mathematician in Residence (DMiR) program. SRG and TS supported by a David L. Hirsch III and Susan H. Hirsch Research Initiation Grant. FL was supported in part by grant CPRR160325161141 and an A-rated researcher award both from the NRF of South Africa and by grant No. 17-02804S of the Czech Granting Agency.

* Corresponding author.
E-mail addresses: stephan.garcia@pomona.edu (S.R. Garcia), Florian.Luca@wits.ac.za (F. Luca), timothyschaaff@gmail.com (T. Schaaff).
URL: http://pages.pomona.edu/~sg064747 (S.R. Garcia).

https://doi.org/10.1016/j.jnt.2017.10.023

Bateman–Horn conjecture
Twin prime conjecture
Brun sieve
Cousin prime
Sexy prime
Prime bias

1. Introduction

Let k be a positive even integer and suppose that p and $p+k$ are prime. Then the difference between the number of primitive roots modulo p and modulo $p+k$ is

$$T(p) := \varphi(p-1) - \varphi(p+k-1).$$

If $T(p) > 0$, then p has more primitive roots than $p+k$ does; if $T(p) < 0$, then p has fewer primitive roots than $p+k$ does. We are interested here in the sign of $T(p)$ as p ranges over the set of all primes p for which $p+k$ is also prime.

To streamline our presentation, we let $\mathbb{P}_k$ denote the set of primes p for which $p+k$ is prime. For example, $\mathbb{P}_2$ is the set of twin primes, $\mathbb{P}_4$ is the set of cousin primes, and $\mathbb{P}_6$ is the set of "sexy primes." We denote by $\mathbb{P}_k(x)$ the set of elements in $\mathbb{P}_k$ that are at most x. The number of elements in $\mathbb{P}_k(x)$ is denoted by $\pi_k(x)$; this is the counting function of $\mathbb{P}_k$. That is, $\pi_k(x)$ is the number of primes $p \leqslant x$ such that $p+k$ is prime. In what follows, the letters p, q, r, s are reserved for primes.

It has long been conjectured that each $\mathbb{P}_k$ is infinite (this appears to date back at least to de Polignac). For example, the *twin-prime conjecture* asserts that $\mathbb{P}_2$ is infinite. There have been tantalizing steps toward this conjecture in recent years [2,5,11]. A more refined version of the twin-prime conjecture is the first *Hardy–Littlewood conjecture*, which asserts that π_2 is asymptotic to a certain constant times $x/(\log x)^2$. The far-reaching *Bateman–Horn conjecture* (Section 2) implies that each $\mathbb{P}_k$ is infinite and provides asymptotics for π_k on the order of $x/(\log x)^2$. The first Hardy–Littlewood conjecture and the twin-prime conjecture both follow from the Bateman–Horn conjecture.

Our work is inspired by [3], in which a peculiar primitive root bias was discovered in the twin prime case $k = 2$. Assuming the Bateman–Horn conjecture, it was proved that at least 65.13% of twin prime pairs $p, p+2$ satisfy $T(p) > 0$ and that at least 0.459% satisfy $T(p) < 0$ (numerical evidence suggests that the bias is approximately 98% to 2%). This is interesting for two reasons. First, a pronounced bias in favor of $T(p) > 0$ exists for twin primes (although relatively easy to motivate from an heuristic standpoint, the proof is long and involved). Second, the bias is not total: the inequality is reversed for a small positive proportion of the twin primes.

In this paper, we extend the results of [3] to prime pairs $p, p+k$. As before, we assume the Bateman–Horn conjecture. Although there are some similarities, many significant complications arise when passing from the case $k = 2$ to $k \geqslant 4$.

(a) The direction and magnitude of the bias in $T(p)$ now depend heavily on the value of $k \pmod 3$ and the smallest primes that do not divide k. If $k \equiv -1 \pmod 3$, then an overwhelming majority of primes $p \in \mathbb{P}_k$ satisfy $T(p) > 0$. If $k \equiv 1 \pmod 3$, then the bias is strongly toward $T(p) < 0$. If $k \equiv 0 \pmod 3$, then the extreme bias disappears and either sign can be favored.

(b) An elementary lemma in the twin-prime case [3, Lem. 2] that relates the sign of $T(p)$ to the sign of a more tractable function fails for $k \geqslant 4$ and must be replaced by a much more difficult asymptotic version (Theorem 6).

(c) The "influence" of the small primes 5, 7, and 11 was sufficient to establish that a positive proportion of twin prime pairs $p, p+2$ satisfy $T(p) < 0$ [3]. This straightforward analysis is no longer possible for $k \geqslant 4$ and we must introduce several parameters in order to compensate.

(d) The tolerances are spectacularly small for certain k. A notable example is $k = 14$. Among the first 20 million primes there are 1,703,216 pairs of primes of the form $p, p+14$; see Table 1. Only three pairs satisfy $T(p) \leqslant 0$, a proportion of 1.76×10^{-6}. These sorts of numbers give us little room to maneuver.

A more extreme example is $k = 70$. Among the first 20 million primes, every prime pair $p, p + 70$ satisfies $T(p) < 0$. Nevertheless, our approach proves that a tiny positive proportion (at least 1.81×10^{-20}) of the primes in $\mathbb{P}_{70}$ satisfy $T(p) > 0$. Even in such lopsided cases, we are able to prove that the biases are not total: the dominant inequality is reversed for a positive proportion of the primes considered.

This paper is organized as follows. Section 2 introduces the Bateman–Horn conjecture and a closely-related unconditional result that is necessary for our work. Section 3 concerns a "totient comparison theorem" (Theorem 6) that permits us to consider a more convenient function $S(p)$ in place of $T(p)$. The short Section 4 contains an heuristic argument that explains the dependence of our results upon the value of $k \pmod 3$. For $k \not\equiv 0 \pmod 3$, the heuristic argument is turned into a rigorous, quantitative theorem in Section 5, which contains our main result (Theorem 7). Although it is too technical to state here, Theorem 7 proves the following.

(a) For $k \not\equiv 0 \pmod 3$, strong primitive root biases exist for prime pairs $p, p + k$.

(b) The biases are not total: the dominant inequality is reversed for a positive proportion of prime pairs $p, p + k$.

We conclude in Section 6 with an analogous theorem (Theorem 10) for $k \equiv 0 \pmod 3$. In this case, we prove that substantial positive proportions of $p \in \mathbb{P}_k$ satisfy $T(p) > 0$ and $T(p) < 0$, respectively. Thus, the extreme biases observed in the $k \not\equiv 0 \pmod 3$ setting disappear.

Table 1
The proportion of prime pairs $p, p+k$ among the first 20 million primes for which p has fewer primitive roots than $p+k$ does. Extreme biases occur for $k \not\equiv 0 \pmod 3$ (see Theorem 7); the situation is more balanced if $k \equiv 0 \pmod 3$ (see Theorem 10).

k	$\#T(p) < 0$	$\pi_k(x)$	Proportion	k	$\#T(p) < 0$	$\pi_k(x)$	Proportion
2	28490	1418478	0.0200849	62	1980	1468111	0.00134867
4	1390701	1419044	0.980027	64	1416847	1418937	0.998527
6	1687207	2836640	0.594791	66	2187908	3153911	0.693713
8	28771	1417738	0.0202936	68	25409	1512639	0.0167978
10	1891800	1891902	0.999946	70	2270424	2270424	1
12	1441259	2837946	0.507853	72	1431789	2837200	0.504649
14	3	1703216	1.76×10^{-6}	74	64	1459313	4.39×10^{-5}
16	1420209	1420273	0.999955	76	1502310	1502338	0.999981
18	1433488	2837906	0.505122	78	1745211	3096187	0.563665
20	4	1891034	2.12×10^{-6}	80	113	1892585	5.95×10^{-5}
22	1576076	1576379	0.999808	82	1426536	1455721	0.979952
24	1015032	2838360	0.357612	84	1145652	3404217	0.336539
26	26521	1546675	0.0171471	86	28787	1454174	0.0197961
28	1699783	1702838	0.998206	88	1553144	1576531	0.985166
30	1930480	3784105	0.510155	90	1489160	3785003	0.393437
32	20553	1418579	0.0144884	92	29413	1486659	0.0197846
34	1495332	1513933	0.987713	94	1421558	1450180	0.980263
36	2097416	2838465	0.738926	96	1915769	2839516	0.674682
38	21739	1502517	0.0144684	98	377	1702580	0.000221429
40	1891651	1891659	0.999996	100	1891334	1891337	0.999998
42	1727098	3405081	0.507212	102	1531067	3027395	0.505737
44	6	1576157	3.81×10^{-6}	104	9	1549054	5.81×10^{-6}
46	1486910	1486946	0.999976	106	1447486	1447486	1
48	1318068	2838746	0.464313	108	1434316	2838777	0.505258
50	48	1891847	2.54×10^{-5}	110	16	2101919	7.61×10^{-6}
52	1525943	1548356	0.985525	112	1699877	1702796	0.998286
54	933772	2839928	0.328801	114	1051285	3004570	0.349895
56	2272	1701628	0.00133519	116	22762	1471017	0.0154736
58	1447184	1472758	0.982635	118	1418455	1442208	0.98353
60	1939665	3783957	0.512602	120	2269102	3784749	0.599538

2. The Bateman–Horn conjecture and Brun's sieve

Let $f_1, f_2, \ldots, f_m$ be a collection of distinct irreducible polynomials with integer coefficients and positive leading coefficients. An integer n is *prime generating* for this collection if each $f_1(n), f_2(n), \ldots, f_m(n)$ is prime. Let $P(x)$ denote the number of prime-generating integers at most x and suppose that $f = f_1 f_2 \cdots f_m$ does not vanish identically modulo any prime. The *Bateman–Horn conjecture* asserts that

$$P(x) \sim \frac{C}{D} \int_2^x \frac{dt}{(\log t)^m},$$

in which

$$D = \prod_{i=1}^m \deg f_i \quad \text{and} \quad C = \prod_p \frac{1 - N_f(p)/p}{(1 - 1/p)^m},$$

Table 2
Numerical approximations of the Bateman–Horn constant C_k.

k	C_k	k	C_k	k	C_k	k	C_k
2	1.32032	32	1.32032	62	1.36585	92	1.3832
4	1.32032	34	1.40835	64	1.32032	94	1.34966
6	2.64065	36	2.64065	66	2.93405	96	2.64065
8	1.32032	38	1.39799	68	1.40835	98	1.58439
10	1.76043	40	1.76043	70	2.11252	100	1.76043
12	2.64065	42	3.16878	72	2.64065	102	2.81669
14	1.58439	44	1.46703	74	1.35805	104	1.44035
16	1.32032	46	1.3832	76	1.39799	106	1.34621
18	2.64065	48	2.64065	78	2.88071	108	2.64065
20	1.76043	50	1.76043	80	1.76043	110	1.95604
22	1.46703	52	1.44035	82	1.35418	112	1.58439
24	2.64065	54	2.64065	84	3.16878	114	2.79598
26	1.44035	56	1.58439	86	1.35253	116	1.36922
28	1.58439	58	1.36922	88	1.46703	118	1.34349
30	3.52086	60	3.52086	90	3.52086	120	3.52086

where $N_f(p)$ is the number of solutions to $f(n) \equiv 0 \pmod{p}$ [1]. For simplicity, we prefer the asymptotically equivalent expression

$$\frac{Cx}{D(\log x)^m}.$$

For a fixed k, let

$$f(t) = t(t+k), \tag{2.1}$$

so that

$$N_f(p) = \begin{cases} 1 & \text{if } p|k, \\ 2 & \text{if } p \nmid k. \end{cases} \tag{2.2}$$

The Bateman–Horn conjecture predicts that

$$\pi_k(x) \sim \prod_{p|k} \frac{p(p-1)}{(p-1)^2} \prod_{p\nmid k} \frac{p(p-2)}{(p-1)^2} \frac{x}{(\log x)^2} = \frac{C_k x}{(\log x)^2},$$

in which

$$C_k = \prod_{p|k} \frac{p(p-1)}{(p-1)^2} \prod_{p\nmid k} \frac{p(p-2)}{(p-1)^2}$$

depends only on the primes that divide k; see Table 2. For example, $C_k \approx 1.32032$ whenever k is a power of 2. In particular, $C_k/2 \approx 0.660162$ is the *twin-primes constant*.

Although weaker than the Bateman–Horn conjecture, the Brun sieve [7, Thm. 3, Sect. I.4.2] suffices for many applications. It does, however, have the distinct advantage of being a proven fact, rather than a long-standing conjecture. The Brun sieve implies that there is a constant B that depends only on m and D such that

$$P(x) \leqslant \frac{BC}{D} \int_2^x \frac{dt}{(\log t)^m} = (1+o(1)) \frac{BC}{D} \frac{x}{(\log x)^m}$$

for sufficiently large x. In particular, there is a constant K such that

$$\pi_k(x) \leqslant K \frac{C_k x}{(\log x)^2}$$

for all k and sufficiently large x. Thus, the Brun sieve implies that the upper bound on π_k implied by the Bateman–Horn conjecture is of the correct order of magnitude.

3. Totient comparison theorem

The well-known formula

$$\frac{\varphi(n)}{n} = \prod_{q|n} \left(1 - \frac{1}{q}\right) \tag{3.1}$$

depends only on the primes that divide n and not on their multiplicity. Because of this, we find it more convenient to work with

$$S(p) := \frac{\varphi(p-1)}{p-1} - \frac{\varphi(p+k-1)}{p+k-1}$$

instead of the more obvious quantity

$$T(p) = \varphi(p-1) - \varphi(p+k-1).$$

We are able to do this because the sign of $T(p)$ almost always agrees with the sign of $S(p)$. For $k = 2$, elementary considerations confirm that $S(p)T(p) > 0$ for $p \geqslant 5$ [3, Lem. 2]. For $k \geqslant 4$, the things are more difficult. We require several lemmas before we obtain an asymptotic analogue of the desired result (Theorem 6).

We first need to estimate the number of $p \in \mathbb{P}_k(x)$ for which $S(p)$ or $T(p)$ equals zero. In both cases, the number is negligible when compared with $\pi_k(x)$; this is Lemma 3 below. To this end, we need the following result.

Lemma 1 *(Graham–Holt–Pomerance [4]). Suppose that j and $j+k$ have the same prime factors. Let $g = \gcd(j, j+k)$ and suppose that*

$$\frac{jt}{g} + 1 \qquad \text{and} \qquad \frac{(j+k)t}{g} + 1 \tag{3.2}$$

are primes that do not divide j.

(a) *Then* $n = j\left(\frac{(j+k)t}{g} + 1\right)$ *satisfies* $\varphi(n) = \varphi(n+k)$.

(b) *For* k *fixed and sufficiently large* x, *the number of solutions* $n \leqslant x$ *to* $\varphi(n) = \varphi(n+k)$ *that are not of the form above is less than* $x/\exp((\log x)^{1/3})$.

Part (b) of the preceding was improved by Yamada [10], although the bound there is slightly more complicated than that of Graham–Holt–Pomerance. In Lemma 1, one considers numbers with the same prime factors. Because of this, we will also need the following lemma of Thue.

Lemma 2 *(Thue [8]). Let* $1 = n_1 < n_2 < \cdots$ *be the sequence of positive integers whose prime factors are at most* p. *Then* $\lim_{i\to\infty}(n_{i+1} - n_i) = \infty$.

A more explicit version of Thue's theorem is due to Tijdeman [9], who proved that there is an effectively computable constant $C = C(p)$ such that $n_{i+1} - n_i > n_i/(\log n_i)^C$ for $n_i \geqslant 3$. For our purposes, however, Thue's result is sufficient. In particular, Lemma 2 implies that for each fixed k, the sequence $n_1, n_2, \ldots$ contains only finitely many pairs n_i, n_j for which $n_j = n_i + k$.

We are now ready to show that $T(p)$ and $S(p)$ are rarely equal to zero relative to the counting function π_k.

Lemma 3. *As* $x \to \infty$,

(a) $\#\{p \in \mathbb{P}_k(x) : S(p) = 0\} = o(\pi_k(x))$, *and*

(b) $\#\{p \in \mathbb{P}_k(x) : T(p) = 0\} = o(\pi_k(x))$.

Proof. (a) Let $P(n)$ denote the largest prime factor of n. Since

$$\frac{\varphi(n)}{n} = \prod_{q|n}\left(\frac{q-1}{q}\right), \tag{3.3}$$

it follows that $P(n)$ is the largest prime factor of the denominator of $\varphi(n)/n$. If $S(p) = 0$, then $P(p-1) = P(p+k-1)$ divides $\gcd(p-1, p+k-1)$, which divides k. Consequently, $S(p) = 0$ implies that the prime factors of both $p-1$ and $p+k-1$ are at most k. Lemma 2 implies that only finitely many such p exist. Thus, the number of primes $p \in \mathbb{P}_k(x)$ for which $S(p) = 0$ is $o(\pi_k(x))$.

(b) Lemma 2 ensures that for each fixed k, there are only finitely many j for which j and $j + k$ have the same prime factors. Fix j and let $g = \gcd(j, j + k)$. To apply Lemma 1 with $n = p - 1$, we must count those

$$t \leqslant \frac{g(x - j + 1)}{j(j + k)} \qquad \text{(so that } p \leqslant x)$$

for which

$$p = j\left(\frac{j+k}{g}t+1\right)+1, \qquad q = j\left(\frac{j+k}{g}t+1\right)+k+1,$$
$$r = \frac{j}{g}t = 1, \qquad s = \frac{j+k}{g}t+1,$$

are simultaneously prime. Since we have four linear constraints, the Brun sieve ensures that the number of such t is $O(x/(\log x)^4) = o(\pi_k(x))$. Thus, the number of primes $p \in \mathbb{P}_k(x)$ for which $T(p) = 0$ is $o(\pi_k(x))$. □

Our proof of Lemma 3(a) actually shows something stronger: $S(p) = 0$ for only finitely many $p \in \mathbb{P}_k$. We can prove Lemma 3(a) as stated without Thue's result (Lemma 2) as follows. If $p - 1 \leqslant x$ and $P(p-1) \leqslant k$, then $p - 1$ is divisible only by the $\pi(k)$ primes at most k. The number of such p at most x is[1] $O((\log x)^{\pi(k)}) = o(x/(\log x)^2)$, even without the condition that p is prime.

The next step toward the desired totient comparison theorem (Theorem 6) is to prove that for each $\ell \geqslant 1$, most $p \in \mathbb{P}_k(x)$ have the property that $2^\ell | T(p)$; this is Lemma 5. Since $T(p) = 0$ rarely occurs by Lemma 3(b), it will follow that $T(p)$ is typically large in absolute value. To do this, we require the following folk lemma. Since we are unable to locate an exact reference for it, we provide the proof.

Lemma 4. $\displaystyle\sum_{q^a \leqslant x} \frac{1}{q^a} = \log\log x + O(1).$

Proof. Mertens' theorem [6, §VII.28.1b] implies that

$$\sum_{q \leqslant x} \frac{1}{q} = \log\log x + O(1).$$

Thus,

$$\begin{aligned}\sum_{q^a \leqslant x} \frac{1}{q^a} &= \sum_{q \leqslant x} \frac{1}{q} + \sum_{\substack{q^a \leqslant x \\ a \geqslant 2}} \frac{1}{q^a} \leqslant \log\log x + O(1) + \sum_{n \geqslant 2}\sum_{k \geqslant 2} \frac{1}{n^k} \\ &= \log\log x + O(1) + \sum_{n \geqslant 2} \frac{1}{n^2} \cdot \frac{1}{1 - 1/n} \\ &= \log\log x + O(1). \qquad \square\end{aligned}$$

[1] If $\pi(k) = s$ and $P(p-1) \leqslant k$, we may write $p - 1 = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}$, in which $2 = p_1 < p_2 < \cdots < p_s$ are the primes at most k. For $i = 1, 2, \ldots, s$, we have $p_i^{a_i} \leqslant x$ and hence $a_i \leqslant (\log x)/\log p_i$. Thus, there are at most $1 + \log x/\log p_i$ possibilities for a_i. Consequently, there are at most $O((\log x + 1)^s) = O((\log x)^{\pi(k)})$ admissible vectors of exponents $(a_1, a_2, \ldots, a_s)$.

Let $\omega(n)$ denote the number of distinct prime divisors of n. The formula

$$\varphi(n) = \prod_{p^a \| n} p^{a-1}(p-1)$$

ensures that $2^{\omega(n)-1} | \varphi(n)$ because each odd prime power p^a that exactly divides n provides at least an additional factor of 2 to $\varphi(n)$ since $p-1$ is even. If p is large, then $p-1$ and $p+k-1$ tend to have many prime factors. Thus, we expect that $T(p)$ should be divisible by a large power of 2. The following makes this precise.

Lemma 5. *For $k \geqslant 2$ even and $\ell \geqslant 1$,*

$$\#\big\{p \in \mathbb{P}_k(x) \,:\, 2^\ell | T(p)\big\} \sim \pi_k(x).$$

Proof. It suffices to show that the counting function for the set of $p \in \mathbb{P}_k(x)$ for which $\omega(p-1) \leqslant \ell$ or $\omega(p+k-1) \leqslant \ell$ is $o(\pi_k(x))$. Indeed, if $\omega(p-1), \omega(p+k-1) \geqslant \ell+1$, then the preceding discussion implies that 2^ℓ divides both $\varphi(p-1)$ and $\varphi(p+k-1)$, and hence divides $T(p)$.

If $\omega(p-1) \leqslant \ell$, then $p-1 = nr$, in which r is prime and $\omega(n) \leqslant \ell$. We must have $\gcd(n, k+1) = 1$ since otherwise $p+k$ would be composite. Let g be the product of the three polynomials

$$g_1(t) = t, \qquad g_2(t) = nt+1, \qquad g_3(t) = nt+k+1.$$

Then

$$N_g(q) = \begin{cases} 1 & \text{if } q|n, \\ 2 & \text{if } q \nmid n \text{ and } q|k \text{ or } q|(k+1), \\ 3 & \text{if } q \nmid n,\ q \nmid k,\ q \nmid (k+1). \end{cases}$$

The Brun sieve provides the following asymptotic estimate, uniformly in n:

$$\begin{aligned}
\sum_{\substack{t \leqslant \frac{x}{n} \\ t,nt+1, \\ nt+k+1 \text{ prime}}} 1 &\ll \frac{x/n}{(\log x/n)^3} \prod_q \frac{1 - N_g(q)/q}{(1-1/q)^3} \\
&\ll \frac{x}{n(\log x)^3} \prod_{q|n} \frac{1-1/q}{(1-1/q)^3} \prod_{\substack{q \nmid n \text{ and} \\ q|k \text{ or } q|(k+1)}} \frac{1-2/q}{(1-1/q)^3} \prod_{\substack{q \nmid n, q \nmid k \\ \text{and } q \nmid (k+1)}} \frac{1-3/q}{(1-1/q)^3} \\
&\ll \frac{x}{n(\log x)^3} \left[\frac{1}{(1-\frac{1}{2})^2}\right]^{\omega(n)} \prod_{\substack{q|k \text{ or} \\ q|(k+1)}} \frac{1-2/q}{(1-1/q)^3} \prod_{\substack{q \nmid n, q \nmid k \\ \text{and } q \nmid (k+1)}} \frac{1-3/q}{(1-1/q)^3} \qquad (3.4) \\
&\ll \frac{2^{2\ell} x}{n(\log x)^3}.
\end{aligned}$$

In the preceding computation, we used the fact that

$$1 \leqslant \frac{1-2/q}{(1-1/q)^3} \quad \text{for } q \geqslant 3$$

to overestimate the finite product in the middle of (3.4) independently of n. Moreover, the third product in (3.4) converges since

$$1 - \frac{1-3/q}{(1-1/q)^3} = \frac{3q-1}{(q-1)^3} \sim \frac{3}{q^2}.$$

Lemma 4 provides

$$\begin{aligned}
\sum_{\substack{p\in\mathbb{P}_k(x)\\ \omega(p-1)\leqslant\ell}} 1 &= \sum_{\substack{n\leqslant x\\ \omega(n)\leqslant\ell}} \sum_{\substack{t\leqslant \frac{x}{n}\\ t,nt+1,\\ nt+k+1 \text{ prime}}} 1 \\
&\ll \sum_{\substack{n\leqslant x\\ \omega(n)\leqslant\ell}} \frac{2^{2\ell}x}{n(\log x)^3} \\
&\ll \frac{x}{(\log x)^3} \sum_{\substack{n\leqslant x\\ \omega(n)\leqslant\ell}} \frac{1}{n} \\
&\ll \frac{x}{(\log x)^3} \frac{1}{\ell!} \Bigg(1 + \sum_{q^a\leqslant x} \frac{1}{q^a}\Bigg)^{\ell} \\
&\ll \frac{x}{(\log x)^3} \big(\log\log x + O(1)\big)^{\ell} \\
&= o(\pi_k(x)).
\end{aligned}$$

Similarly, the count of $p \in \mathbb{P}_k(x)$ with $\omega(p+k-1) \leqslant \ell$ is also $o(\pi_k(x))$. □

We are now in a position to prove the main result of this section. It says that $S(p)$ and $T(p)$ are nonzero and share the same sign for most $p \in \mathbb{P}_k(x)$.

Theorem 6 *(Totient comparison theorem). Let k be even. Then as $x \to \infty$,*

$$\#\{p \in \mathbb{P}_k(x) : S(p)T(p) > 0\} \sim \pi_k(x).$$

Proof. In light of Lemma 3, it suffices to show that

$$\#\{p \in \mathbb{P}_k(x) : S(p) > 0,\, T(p) > 0\} \sim \pi_k(x).$$

Since $T(p) > 0$ implies that $S(p) > 0$, we focus on the converse. If $S(p) > 0$, then

$$\begin{aligned} 0 &< (p+k-1)\varphi(p-1) - (p-1)\varphi(p+k-1) \\ &= p\big(\varphi(p-1) - \varphi(p+k-1)\big) + (k-1)\varphi(p-1) + \varphi(p+k-1) \\ &\leqslant p\big(\varphi(p-1) - \varphi(p+k-1)\big) + (k-1)(p-1) + (p+k-1) \\ &\leqslant p\big(\varphi(p-1) - \varphi(p+k-1) + k\big) \\ &= p\big(T(p) + k\big). \end{aligned} \tag{3.5}$$

Fix ℓ so that $2^\ell \geqslant k$. Apply Lemma 5 at (3.5) and conclude that

$$\{p \in \mathbb{P}_k(x) : S(p) > 0,\, T(p) \geqslant 0\} \sim \pi_k(x).$$

Now apply Lemma 3(b) to replace $T(p) \geqslant 0$ in the preceding with $T(p) > 0$. □

In light of Theorem 6, we can focus our attention on the expression $S(p)$, which is nonzero and shares the same sign as $T(p)$ for all $p \in \mathbb{P}_k$ outside of a set of zero density with respect to the counting function $\pi_k(x)$. The two expressions

$$\frac{\varphi(p-1)}{p-1} = \prod_{q|(p-1)} \left(1 - \frac{1}{q}\right) \quad \text{and} \quad \frac{\varphi(p+k-1)}{p+k-1} = \prod_{q|(p+k-1)} \left(1 - \frac{1}{q}\right) \tag{3.6}$$

that comprise $S(p)$ are primarily determined by the small prime divisors of $p-1$ and $p+k-1$. Since p and $p+k$ are both prime, the nature of these small divisors is also related to k.

4. An heuristic argument

Before proceeding to the technical details, it is instructive to go through a brief heuristic argument. With the help of the Bateman–Horn conjecture, we will ultimately be able to turn this informal reasoning into rigorous, quantitative proofs.

As Table 1 suggests, the behavior of $T(p)$ is heavily influenced by the value of $k \pmod 3$. Here is the explanation.

- If $k \equiv -1 \pmod 3$, then elementary considerations imply that $3|(p+k-1)$ whenever $p, p+k$ are prime and $p \geqslant 5$. Then (3.6) becomes

$$\frac{\varphi(p-1)}{p-1} = \frac{1}{2} \prod_{\substack{q \geqslant 5 \\ q|(p-1) \\ q\nmid(k+1)}} \left(1 - \frac{1}{q}\right) \quad \text{and} \quad \frac{\varphi(p+k-1)}{p+k-1} = \frac{1}{3} \prod_{\substack{q \geqslant 5 \\ q|(p+k-1) \\ q\nmid(k-1)}} \left(1 - \frac{1}{q}\right),$$

  and hence we expect that $S(p) > 0$ for most $p \in \mathbb{P}_k$. Moreover, this suggests that $S(p) < 0$ might occur if $p-1$ is divisible by many small primes.

- If $k \equiv 1 \pmod 3$, then a similar argument tells us that

$$\frac{\varphi(p-1)}{p-1} = \frac{1}{3} \prod_{\substack{q \geqslant 5 \\ q \mid (p-1) \\ q \nmid (k+1)}} \left(1 - \frac{1}{q}\right) \quad \text{and} \quad \frac{\varphi(p+k-1)}{p+k-1} = \frac{1}{2} \prod_{\substack{q \geqslant 5 \\ q \mid (p+k-1) \\ q \nmid (k-1)}} \left(1 - \frac{1}{q}\right).$$

  Thus, we expect that $S(p) < 0$ for most $p \in \mathbb{P}_k$ and that $S(p) > 0$ might occur if $p+k-1$ is divisible by many small primes.
- If $k \equiv 0 \pmod q$, in which q is prime, then q either divides both $p-1$ and $p+k-1$, or it divides neither. Thus, the prime divisors of k have no bearing upon the large-scale sign behavior of $S(p)$. It is the small primes $q \geqslant 5$ that divide exactly one of $p-1$ and $p+k-1$ which govern our problem. Consequently, the observed bias in the sign of $S(p)$ is less pronounced if $3|k$.

5. Primitive roots biases for $k \not\equiv 0 \pmod 3$

Let χ_3 denote the nontrivial Dirichlet character modulo 3. That is,

$$\chi_3(k) = \begin{cases} 0 & \text{if } k \equiv 0 \pmod 3, \\ 1 & \text{if } k \equiv 1 \pmod 3, \\ -1 & \text{if } k \equiv 2 \pmod 3. \end{cases}$$

Fix $k \not\equiv 0 \pmod 3$ and let

$$5 \leqslant q_1 < q_2 < q_3 < \cdots$$

be the ordered sequence of primes that do not divide

$$k(k - \chi_3(k)), \tag{5.1}$$

which is a multiple of 6. This sequence is infinite since it contains all primes larger than $\max\{5, k+1\}$. Let $Q = Q(k)$ denote the set

$$Q = \{q_1, q_2, \ldots, q_m\}, \tag{5.2}$$

in which the index m shall be determined momentarily. Define

$$L_k = \log\left[\frac{2}{3} \prod_{q \in Q} \left(1 + \frac{1}{q-1}\right)\right] \tag{5.3}$$

and

$$R_k = \sum_{\substack{r \geqslant 5 \\ r \notin Q \\ r \nmid (k+\chi_3(k))}} \frac{1}{r - N_f(r)} \log\left(1 + \frac{1}{r-1}\right), \tag{5.4}$$

in which $f(t) = t(t+k)$ is the polynomial defined in (2.1). From (2.2), we see that

$$r - N_f(r) \in \{r-1, r-2\}$$

for all primes r, so the general term in (5.4) is $O(1/r^2)$. Define m in (5.2) to be the smallest index such that

$$L_k > R_k. \tag{5.5}$$

This is possible since the product (5.3) diverges if taken over all sufficiently large primes, while the sum (5.4) converges under the same circumstances.

This establishes the notation necessary for part (a) of the following result. For part (b), we use an expression similar to (5.4). Let

$$R'_k = \sum_{\substack{r \geqslant 5 \\ r \nmid (k-\chi_3(k))}} \frac{1}{r - N_f(r)} \log\left(1 + \frac{1}{r-1}\right).$$

This lays the foundation for the following theorem, which establishes a bias in the number of primitive roots of prime pairs $p, p+k$ when $k \equiv \pm 1 \pmod 3$.

Theorem 7. *Assume that the Bateman–Horn conjecture holds. Let $k \not\equiv 0 \pmod 3$.*

(a) *The set of primes $p \in \mathbb{P}_k$ for which*

$$\operatorname{sgn} T(p) = \chi_3(k)$$

has lower density (as a subset of $\mathbb{P}_k$) at least

$$\prod_{q \in Q} (q-2)^{-1} \left(1 - \frac{R_k}{L_k}\right) > 0.$$

(b) *The set of primes $p \in \mathbb{P}_k$ for which*

$$\operatorname{sgn} T(p) = -\chi_3(k)$$

has lower density (as a subset of all prime pairs $p, p+k$) at least

$$1 - \frac{R'_k}{\log(3/2)} > 0.6515.$$

Table 3
Lower bounds in Theorem 7 for $k \equiv -1 \pmod 3$.

k	Q	L_k	R_k	Lower bound: $T(p) < 0$	R'_k	Lower bound: $T(p) > 0$
2	5, 7, 11	0.067139	0.025497	0.004594	0.141298	0.651516
8	5, 7, 11	0.067139	0.025497	0.004594	0.141298	0.651516
14	11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53	0.113089	0.103683	1.56×10^{-18}	0.061779	0.847635
20	11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47	0.094041	0.090599	3.50×10^{-17}	0.091873	0.773414
26	5, 7, 11	0.067139	0.024890	0.004661	0.140692	0.653012
32	5, 7, 13	0.051872	0.027680	0.002826	0.130708	0.677634
38	5, 7, 11	0.067139	0.0245373	0.004700	0.133845	0.669898
44	7, 13, 17, 19, 23, 29, 31, 37, 41	0.107845	0.088373	5.73×10^{-13}	0.065858	0.837574
50	7, 11, 13, 19, 23, 29, 31	0.090439	0.066279	1.93×10^{-9}	0.118661	0.707346
56	5, 11, 13, 17	0.053656	0.039870	0.000057	0.132979	0.672033
62	5, 11, 13, 17	0.053656	0.044691	0.000037	0.11043	0.727645
68	5, 7, 11	0.067140	0.025013	0.004647	0.138929	0.65736
74	7, 11, 13, 17, 19, 23	0.083182	0.083047	6.14×10^{-10}	0.066895	0.835016
80	7, 11, 13, 17, 19, 23	0.083182	0.064502	8.47×10^{-8}	0.122703	0.697378
86	5, 7, 11	0.067139	0.0214415	0.005041	0.139985	0.654755
92	5, 7, 11	0.067139	0.018124	0.005407	0.140071	0.654542
98	5, 13, 17, 19, 23	0.056865	0.045054	1.17×10^{-6}	0.125570	0.690307
104	11, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79	0.122425	0.114018	1.71×10^{-28}	0.035480	0.912495
110	7, 13, 17, 19, 23, 29, 31, 41	0.080446	0.071049	1.29×10^{-11}	0.120861	0.701920
116	5, 7, 11	0.067139	0.023334	0.004833	0.133975	0.669577
122	5, 7, 11	0.067139	0.025492	0.004594	0.140660	0.653089
128	5, 7, 11	0.067139	0.025434	0.004601	0.140724	0.652931
134	7, 11, 13, 17, 19, 23, 29	0.118273	0.081959	4.29×10^{-9}	0.066913	0.834971
140	11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 53	0.091583	0.085513	5.59×10^{-17}	0.117087	0.711229
146	5, 11, 13, 17	0.053656	0.043706	0.000041	0.110465	0.727559
152	5, 7, 11	0.067139	0.025276	0.004618	0.137080	0.661920
158	5, 7, 11	0.067139	0.025453	0.004599	0.140922	0.652442
164	7, 13, 17, 19, 23, 29, 31, 37, 43	0.106682	0.090011	4.72×10^{-13}	0.056311	0.861120

Tables 3 and 4 provide the sets Q, numerical values for L_k, R_k, R'_k, and the bounds in Theorem 7 for various values of k.

5.1. Preliminary lemmas

Before proceeding with the proof of Theorem 7, we require a few preliminary results. Certain conditions in Lemmas 8 and 9 are slightly more general than necessary. This is because they will later be applied when $k \equiv 0 \pmod 3$ (Section 6). For our present

Table 4
Lower bounds in Theorem 7 for $k \equiv 1 \pmod 3$.

k	Q	L_k	R_k	Lower bound: $T(p) > 0$	R'_k	Lower bound: $T(p) < 0$
4	5, 7, 11	0.067139	0.025497	0.004594	0.141298	0.651516
10	7, 11, 13, 17, 19, 23	0.083182	0.064667	8.39×10^{-8}	0.122703	0.697378
16	7, 11, 13, 17, 19, 23, 29	0.118273	0.081963	4.28×10^{-9}	0.066917	0.834963
22	5, 13, 17, 19, 23	0.056865	0.049243	7.58×10^{-7}	0.109409	0.730164
28	5, 11, 13, 17	0.053656	0.038571	0.000063	0.13616	0.664189
34	5, 7, 13	0.051872	0.028558	0.002723	0.130455	0.678257
40	7, 11, 17, 19, 23, 29, 31	0.071021	0.068880	1.59×10^{-10}	0.115426	0.715324
46	7, 11, 13, 17, 19, 29, 31	0.106611	0.082375	2.30×10^{-9}	0.066821	0.8352
52	5, 7, 11	0.067139	0.024517	0.004702	0.13665	0.662979
58	5, 7, 11	0.067139	0.025150	0.004632	0.138071	0.659474
64	5, 11, 13, 17	0.053656	0.045009	0.000036	0.110468	0.727552
70	11, 13, 17, 19, 29, 31, 37, 41, 43, 47, 53, 59, 61	0.102261	0.086419	1.81×10^{-20}	0.115448	0.715271
76	7, 11, 13, 17, 23, 29, 31	0.096996	0.083859	1.11×10^{-9}	0.066740	0.835398
82	5, 7, 11	0.067139	0.025331	0.004612	0.141282	0.651555
88	5, 7, 13	0.051872	0.027621	0.002833	0.138939	0.657333
94	5, 7, 11	0.067139	0.022306	0.004946	0.140157	0.65433
100	7, 13, 17, 19, 23, 29, 31, 37	0.083152	0.071944	1.67×10^{-11}	0.112113	0.723496
106	11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 59	0.111135	0.108798	3.52×10^{-19}	0.036080	0.911017
112	5, 11, 13, 17	0.053656	0.039790	0.000058	0.135377	0.666119
118	5, 7, 11	0.067139	0.02145	0.005040	0.134016	0.669475
124	5, 7, 11	0.067139	0.025459	0.004599	0.140627	0.65317
130	7, 11, 17, 19, 23, 29, 31	0.071021	0.068848	1.62×10^{-10}	0.121523	0.700289
136	7, 11, 13, 19, 23, 29, 31	0.090439	0.084567	4.69×10^{-10}	0.066664	0.835586
142	5, 7, 11	0.067139	0.018217	0.005397	0.140817	0.652702
148	5, 11, 13, 17	0.053656	0.044941	0.000036	0.110446	0.727606
154	5, 13, 19, 23, 29, 31	0.064121	0.045715	3.11×10^{-8}	0.131059	0.676768
160	7, 11, 13, 17, 19, 23	0.083182	0.064667	8.39×10^{-8}	0.122329	0.698299
166	7, 13, 17, 19, 23, 29, 31, 37, 41	0.107845	0.089968	5.26×10^{-13}	0.056325	0.861085
172	5, 7, 11	0.067139	0.025449	0.004599	0.138104	0.659394
178	5, 7, 11	0.067139	0.025464	0.004598	0.140997	0.652259
184	5, 7, 11	0.067139	0.024618	0.004691	0.140922	0.652444

purposes (the proof of Theorem 7), the set Q in the following lemmas is as defined in the preceding section.

Lemma 8. *Assume that the Bateman–Horn conjecture holds. Let k be a positive even integer and let Q be a finite set of primes such that*

$$q \nmid k(k+1) \qquad (\textit{resp.}, \ q \nmid k(k-1)),$$

for all $q \in Q$. *The number of* $p \in \mathbb{P}_k(x)$ *such that*

$$q|(p-1) \qquad (\textit{resp.}, \ q|(p+k-1))$$

for all $q \in Q$ *is*

$$\pi_k'(x) \ = \ (1+o(1))\, \pi_k(x) \prod_{q \in Q} (q-2)^{-1}.$$

Proof. Suppose that $q \nmid k(k+1)$ for all $q \in Q$, since the case $q \nmid k(k-1)$ is analogous. We wish to count the number of $p \in \mathbb{P}_k(x)$ such that $q|(p-1)$ for all $q \in Q$. If $a = \prod_{q \in Q} q$, then the desired primes are those of the form

$$n = at + 1 \leqslant x \quad \text{such that} \quad n + k = at + k + 1 \text{ is prime.}$$

Let

$$g_1(t) = at + 1, \quad g_2(t) = at + k + 1, \quad \text{and } g = g_1 g_2.$$

In the Bateman–Horn conjecture with s denoting an arbitrary prime, we have

$$N_g(s) = \begin{cases} 0 & \text{if } s \in Q, \\ 1 & \text{if } s|k, \ s \notin Q, \\ 2 & \text{if } s \nmid k, \ \ s \notin Q. \end{cases} \tag{5.6}$$

For sufficiently large x, the Bateman–Horn conjecture predicts that the number of such $t \leqslant (x-1)/a$ is

$$\begin{aligned} \pi_k'(x) &= (1+o(1)) \frac{(x-1)/a}{(\log((x-1)/a))^2} \prod_{s \geqslant 2} \frac{1 - N_g(s)/s}{(1-1/s)^2} \\ &= (1+o(1)) \frac{x}{a(\log x)^2} \prod_{s \geqslant 2} \frac{1 - N_g(s)/s}{(1-1/s)^2} \\ &= (1+o(1)) \frac{x}{a(\log x)^2} \prod_{s \in Q} \frac{1}{(1-1/s)^2} \prod_{s \notin Q} \frac{1 - N_g(s)/s}{(1-1/s)^2} \\ &= (1+o(1)) \frac{x}{a(\log x)^2} \prod_{s \in Q} (1 - N_f(s)/s)^{-1} \prod_{s \geqslant 2} \frac{1 - N_f(s)/s}{(1-1/s)^2}, \end{aligned}$$

in which $N_f(s)$ refers to (2.2). Simplifying further yields

$$\begin{aligned}\pi'_k(x) &= (1+o(1))\frac{\pi_k(x)}{a}\prod_{s\in Q}(1-2/s)^{-1}\\ &= (1+o(1))\pi_k(x)\prod_{q\in Q}(q-2)^{-1}. \qquad \square\end{aligned}$$

Lemma 9. *Assume that the Bateman–Horn conjecture holds. Let k be a positive even integer and let Q be a finite set of primes such that*

$$q \nmid k(k+1) \qquad (\textit{resp.}, \ q\nmid k(k-1)),$$

for all $q \in Q$. Let $r \geqslant 5$ be a fixed prime not in Q that satisfies

$$r \nmid k(k-1) \qquad (\textit{resp.}, \ r\nmid k(k+1)).$$

The number of $p \in \mathbb{P}_k(x)$ such that

$$q|(p-1) \quad \textit{and} \quad r|(p+k-1) \qquad (\textit{resp.}, \ r|(p-1)),$$

for all $q \in Q$ is

$$\pi'_{k,r}(x) = (1+o(1))\,\frac{\pi_k(x)}{r-N_f(r)}\,\prod_{q\in Q}(q-2)^{-1},$$

in which $N_f(r)$ refers to (2.2).

Proof. Suppose that $q \nmid (k+1)$ for all $q \in Q$, since the case $q \nmid (k-1)$ is analogous. Fix a prime $r \geqslant 5$ such that $r \nmid (k-1)$ and let $a = \prod_{q\in Q} q$. The desired primes are precisely those of the form

$$n = aj+1 \leqslant x \quad \text{such that} \quad n+k = aj+k+1 \text{ is prime and } r|(aj+k).$$

In particular, j must be of the form

$$j = j_0 + r\ell,$$

in which j_0 is the smallest positive integer such that $j_0 \equiv -ka^{-1} \pmod{r}$ (note that a is invertible modulo r since $r \notin Q$). Let $b_r = aj_0 + 1$. Then

$$n = ar\ell + b_r \qquad \text{and} \qquad n+k = ar\ell + (b_r+k) \tag{5.7}$$

are both prime, $n \leqslant x$, and

$$\ell \leqslant \frac{x-b_r}{ar}.$$

In the Bateman–Horn conjecture, let

$$g_1(t) = art + b_r, \qquad g_2(t) = art + (b_r + k), \quad \text{and} \quad g = g_1 g_2.$$

With s denoting an arbitrary prime, $N_g(s)$ is as in (5.6) except for $s = r$, in which case $N_g(r) = 0$. Indeed,

$$g_1(t) \equiv b_r \equiv aj_0 + 1 \equiv -k + 1 \not\equiv 0 \pmod r \quad \text{and} \quad g_2(t) \equiv b_r + k \equiv 1 \pmod r$$

for all t. As $x \to \infty$, the Bateman–Horn conjecture predicts that the number of such ℓ is

$$\begin{aligned}
\pi'_{k,r}(x) &= (1 + o(1)) \frac{(x - b_r)/(ar)}{(\log((x - b_r)/(ar)))^2} \prod_{s \geqslant 2} \left(\frac{1 - N_g(s)/s}{(1 - 1/s)^2} \right) \\
&= (1 + o(1)) \frac{x}{ar(\log x)^2} \prod_{s \geqslant 2} \left(\frac{1 - N_g(s)/s}{(1 - 1/s)^2} \right) \\
&= (1 + o(1)) \frac{x}{ar(\log x)^2} \prod_{s \in Q \text{ or } s = r} \left(\frac{1}{(1 - 1/s)^2} \right) \prod_{s \notin Q, s \neq r} \left(\frac{1 - N_g(s)/s}{(1 - 1/s)^2} \right) \\
&= (1 + o(1)) \frac{x}{ar(\log x)^2} \prod_{s \in Q \text{ or } s = r} (1 - N_f(s)/s)^{-1} \prod_{s \geqslant 2} \left(\frac{1 - N_f(s)/s}{(1 - 1/s)^2} \right) \\
&= (1 + o(1)) \frac{\pi_k(x)}{r - N_f(r)} \prod_{q \in Q} (q - 2)^{-1}. \qquad \square
\end{aligned}$$

5.2. Proof of Theorem 7(a)

In light of Theorem 6, we may use $S(p)$ and $T(p)$ interchangeably in what follows. Suppose that $k \not\equiv 0 \pmod 3$.

- If $\chi_3(k) = -1$, then we wish to count $p \in \mathbb{P}_k$ for which $q|(p-1)$ for all $q \in Q$.
- If $\chi_3(k) = 1$, then we wish to count $p \in \mathbb{P}_k$ for which $q|(p+k-1)$ for all $q \in Q$.

Because of this slight difference, we define $\tau_k = k(1 + \chi_3(k))/2$. That is,

$$\tau_k = \begin{cases} 0 & \text{if } \chi_3(k) = -1, \\ k & \text{if } \chi_3(k) = 1, \end{cases} \tag{5.8}$$

so that

$$p - 1 + \tau_k = \begin{cases} p - 1 & \text{if } \chi_3(k) = -1, \\ p + k - 1 & \text{if } \chi_3(k) = 1. \end{cases}$$

Now let $\pi'_k(x)$ denote the number of primes $p \in \mathbb{P}_k(x)$ such that $q|(p-1+\tau_k)$ for all $q \in Q$. Lemma 8 allows us to count these prime pairs. Moreover, Lemma 9 permits us to count such pairs after imposing the additional restriction that a fixed prime $r \geqslant 5$ not in Q divides $p-1+(k-\tau_k)$, where

$$p-1+(k-\tau_k) = \begin{cases} p+k-1 & \text{if } \chi_3(k) = -1, \\ p-1 & \text{if } \chi_3(k) = 1. \end{cases}$$

Let $\pi'_{k,r}(x)$ denote the number of primes $p \in \mathbb{P}_k(x)$ such that $q|(p-1+\tau_k)$ for all $q \in Q$, and $r|(p-1+(k-\tau_k))$.

Suppose that p is counted by $\pi'_k(x)$. The condition $k \not\equiv 0 \pmod{q}$ ensures that $q \nmid (p-1+(k-\tau_k))$ for all $q \in Q$. Thus,

$$3|(p-1+(k-\tau_k)) \qquad \text{and} \qquad 3 \nmid (p-1+\tau_k),$$

so that

$$\frac{\varphi(p-1+\tau_k)}{p-1+\tau_k} \leqslant \frac{1}{2} \prod_{q \in Q} \left(1 - \frac{1}{q}\right).$$

If $\operatorname{sgn} S(p) = -\chi_3(k)$ (so that p does not belong to the set of interest in Theorem 7(a)), then

$$\begin{aligned} \frac{1}{3} \prod_{\substack{r|(p-1+(k-\tau_k)) \\ r\geqslant 5, r\notin Q}} \left(1 - \frac{1}{r}\right) &= \frac{\varphi(p-1+(k-\tau_k))}{p-1+(k-\tau_k)} \\ &< \frac{\varphi(p-1+\tau_k)}{p-1+\tau_k} \leqslant \frac{1}{2} \prod_{q \in Q} \left(1 - \frac{1}{q}\right). \end{aligned}$$

Consequently,

$$\prod_{\substack{r|(p-1+(k-\tau_k)) \\ r\geqslant 5, r\notin Q}} \left(1 + \frac{1}{r-1}\right) > \frac{2}{3} \prod_{q \in Q} \left(1 + \frac{1}{q-1}\right),$$

in which r is prime. Let

$$F(p) := \sum_{\substack{r|(p-1+(k-\tau_k)) \\ r\geqslant 5}} \log\left(1 + \frac{1}{r-1}\right).$$

We want to count primes $p \in \mathbb{P}_k(x)$ such that

$$F(p) > \log\left[\frac{2}{3} \prod_{q \in Q} \left(1 + \frac{1}{q-1}\right)\right] = L_k$$

and $q|(p-1+\tau_k)$ for all $q \in Q$. To do this, we first sum up $F(p)$ over all primes p counted by $\pi'_k(x)$ and change the order of summation to get

$$\begin{aligned} A(x) &= \sum_{\substack{p \text{ counted by} \\ \pi'_k(x)}} F(p) \\ &= \sum_{\substack{r \geqslant 5 \\ r \notin Q}} \pi'_{k,r}(x) \log\left(1 + \frac{1}{r-1}\right) \\ &\leqslant \sum_{\substack{5 \leqslant r \leqslant z \\ r \notin Q}} \pi'_{k,r}(x) \log\left(1 + \frac{1}{r-1}\right) \\ &\qquad + \sum_{\substack{z < r \leqslant (\log x)^3 \\ r \notin Q}} \pi'_{k,r}(x) \log\left(1 + \frac{1}{r-1}\right) \\ &\qquad\qquad + \sum_{\substack{(\log x)^3 < r \leqslant x \\ r \notin Q}} \pi'_{k,r}(x) \log\left(1 + \frac{1}{r-1}\right) \\ &= A_1(x) + A_2(x) + A_3(x), \end{aligned} \tag{5.9}$$

in which z is a fixed number. We bound the three summands separately. In what follows, we let $\delta > 0$ be small, and fix z large enough so that

$$\frac{8K}{z-2} \prod_{q \in Q} (q-2)^{-1} < \frac{\delta}{3}.$$

(a) Suppose that $5 \leqslant r \leqslant z$ and $r \notin Q$. Lemma 9 asserts that if $r \nmid (k + \chi_3(k))$, then

$$\pi'_{k,r}(x) = (1 + o(1)) \frac{\pi_k(x)}{r - N_f(r)} \prod_{q \in Q} (q-2)^{-1}$$

uniformly for $r \in [5, z] \setminus Q$ as $x \to \infty$. If $r|(k + \chi_3(k))$ and $r|(p - 1 + (k - \tau_k))$, then

$$0 \equiv p - 1 + (k - \tau_k) \equiv p - 1 - (\chi_3(k) + \tau_k) \pmod{r}.$$

When $\chi_3(k) = -1$, we have $p \equiv 0 \pmod{r}$. When $\chi_3(k) = 1$, we add $k + \chi_3(k)$ to the middle expression and simplify to get $p + k \equiv 0 \pmod{3}$. In either case, it follows that $\pi'_{k,r}(x) \leqslant 1$ when $r|(k + \chi_3(k))$. Thus, for sufficiently large x we have

$$
\begin{aligned}
A_1(x) &\leqslant (1+o(1))\frac{\pi_k(x)}{\prod_{q\in Q}(q-2)} \sum_{\substack{5\leqslant r\leqslant z\\ r\notin Q, r\nmid(k+\chi_3(k))}} \frac{1}{r-N_f(r)}\log\left(1+\frac{1}{r-1}\right)\\
&\quad + \sum_{\substack{5\leqslant r\leqslant z\\ r\notin Q, r\mid(k+\chi_3(k))}} \log\left(1+\frac{1}{r-1}\right)\\
&\leqslant (1+o(1))\frac{R_k\,\pi_k(x)}{\prod_{q\in Q}(q-2)} + \sum_{\substack{5\leqslant r\leqslant z\\ r\notin Q, r\mid(k+\chi_3(k))}} \log\left(1+\frac{1}{r-1}\right)\\
&= \left((1+o(1))\,\frac{R_k}{\prod_{q\in Q}(q-2)} + \frac{1}{\pi_k(x)} \sum_{\substack{5\leqslant r\leqslant z\\ r\notin Q, r\mid(k+\chi_3(k))}} \log\left(1+\frac{1}{r-1}\right)\right)\pi_k(x)\\
&< \left(\frac{R_k}{\prod_{q\in Q}(q-2)} + \frac{\delta}{3}\right)\pi_k(x),
\end{aligned}
$$

where the last inequality follows from $\pi_k(x) \to \infty$.

(b) Suppose that $z < r \leqslant (\log x)^3$ and $r \notin Q$. Maintaining the notation a, b_r from the proof of Lemma 9, the Brun sieve yields an absolute constant K such that for sufficiently large x,

$$
\begin{aligned}
\pi'_{k,r}(x) &\leqslant \frac{K(x+\tau_k-b_r)/ar}{(\log((x+\tau_k-b_r)/ar))^2}\prod_{p\geqslant 2}\frac{1-N_g(p)/p}{(1-1/p)^2}\\
&= \frac{C_kK(x+\tau_k-b_r)}{(r-N_f(r))(\log((x+\tau_k-b_r)/ar))^2}\prod_{q\in Q}(q-2)^{-1}\\
&\leqslant \frac{2C_kKx}{(r-N_f(r))(\log((x-b_r)/ar))^2}\prod_{q\in Q}(q-2)^{-1}\\
&\leqslant \frac{2C_kKx}{(r-N_f(r))(\log(x/ar-1))^2}\prod_{q\in Q}(q-2)^{-1},
\end{aligned}
$$

where the last inequality follows from the fact that $b_r \leqslant ar$. Since $r \leqslant (\log x)^3$,

$$
\log(x/ar-1) \geqslant \log(x^{1/2}) = (\log x)/2
$$

for large enough x. Thus,

$$
\begin{aligned}
\pi'_{k,r}(x) &\leqslant \frac{8C_kKx}{(r-N_f(r))(\log x)^2}\prod_{q\in Q}(q-2)^{-1}\\
&\leqslant \frac{8K\pi_k(x)}{r-2}\prod_{q\in Q}(q-2)^{-1}
\end{aligned}
$$

for sufficiently large x. Since $\log(1+t) < t$ for $t > 0$, for sufficiently large x we obtain

$$\begin{aligned}
A_2(x) &= \sum_{\substack{z<r\leqslant(\log x)^3\\ r\notin Q}} \pi'_{k,r}(x)\log\left(1+\frac{1}{r-1}\right)\\
&\leqslant \frac{8K\pi_k(x)}{\prod_{q\in Q}(q-2)}\sum_{r>z}\frac{1}{r-2}\log\left(1+\frac{1}{r-1}\right)\\
&< \frac{8K\pi_k(x)}{\prod_{q\in Q}(q-2)}\sum_{r>z}\frac{1}{(r-2)(r-1)}\\
&= \frac{8K\pi_k(x)}{\prod_{q\in Q}(q-2)}\sum_{r>z}\left(\frac{1}{r-2}-\frac{1}{r-1}\right)\\
&\leqslant \frac{8K\pi_k(x)}{(z-2)\prod_{q\in Q}(q-2)}\\
&< \frac{\delta}{3}\,\pi_k(x).
\end{aligned}$$

(c) Suppose that $(\log x)^3 < r \leqslant x$ and $r \notin Q$. By (5.7), the primes counted by $\pi'_{k,r}(x)$ lie in an arithmetic progression modulo ar, with a defined as in Lemma 9. Thus, their number is at most

$$\pi'_{k,r}(x) \leqslant \left\lfloor\frac{x}{ar}\right\rfloor + 1 \leqslant \frac{x}{ar} + 1.$$

Since $\log(1+t) < t$, for sufficiently large x we obtain

$$\begin{aligned}
A_3(x) &= \sum_{\substack{(\log x)^3<r\leqslant x\\ r\notin Q}} \pi'_{k,r}(x)\log\left(1+\frac{1}{r-1}\right)\\
&\leqslant \sum_{\substack{(\log x)^3<r\leqslant x\\ r\notin Q}} \frac{1}{(r-1)}\left(\frac{x}{ar}+1\right)\\
&\leqslant \frac{x}{a}\sum_{r>(\log x)^3}\frac{1}{r(r-1)} + \sum_{(\log x)^3<r\leqslant x}\frac{1}{r-1}\\
&\leqslant \frac{x}{a}\sum_{r>(\log x)^3}\left(\frac{1}{r-1}-\frac{1}{r}\right) + \int_{(\log x)^3-2}^{x}\frac{dt}{t}\\
&\leqslant \frac{x}{a((\log x)^3-1)} + \left(\log t\Big|_{t=(\log x)^3-2}^{t=x}\right)\\
&\leqslant \frac{x}{a((\log x)^3-1)} + \log x
\end{aligned}$$

$$
\begin{aligned}
&= (1+o(1))\left(\frac{x}{a(\log x)^3} + \log x\right)\\
&= o(1)\,\pi_k(x)\\
&< \frac{\delta}{3}\,\pi_k(x).
\end{aligned}
$$

Returning to (5.9) and using the preceding three estimates, we have

$$
\begin{aligned}
A(x) &= A_1(x) + A_2(x) + A_3(x)\\
&< \left(\frac{R_k}{\prod_{q\in Q}(q-2)} + \delta\right)\pi_k(x)
\end{aligned}
$$

for sufficiently large x. Let $\mathcal{U}(x)$ be the set of primes p counted by $\pi_k'(x)$ with $\operatorname{sgn} S(p) = -\chi_3(k)$, so that p does not belong to the set of interest in Theorem 7(a). As we have seen, if $p \in \mathcal{U}(x)$, then

$$
F(p) > L_k.
$$

Thus,

$$
\begin{aligned}
0 &\leqslant \#\mathcal{U}(x)\, L_k\\
&< \sum_{p\in\mathcal{U}(x)} F(p) \leqslant A(x)\\
&< \left(\frac{R_k}{\prod_{q\in Q}(q-2)} + \delta\right)\pi_k(x),
\end{aligned}
$$

from which we deduce that

$$
\#\mathcal{U}(x) < \left(\frac{R_k\prod_{q\in Q}(q-2)^{-1} + \delta}{L_k}\right)\pi_k(x).
$$

The primes p counted by $\pi_k'(x)$ which are not in $\mathcal{U}(x)$ satisfy $\operatorname{sgn} S(p) = \chi_3(k)$. By Lemma 8 and the preceding calculation, for large x there are at least

$$
\begin{aligned}
\pi_k'(x) - \#\mathcal{U}(x) &> \left((1+o(1))\prod_{q\in Q}(q-2)^{-1} - \frac{R_k\prod_{q\in Q}(q-2)^{-1}+\delta}{L_k}\right)\pi_k(x)\\
&= \prod_{q\in Q}(q-2)^{-1}\left(1 - \frac{R_k}{L_k} - \epsilon\right)\pi_k(x)
\end{aligned}
$$

such primes, where $\epsilon > 0$ can be made arbitrarily small by taking x large enough. The condition (5.5) ensures that the quantity in parentheses is positive for a small enough ϵ. By Theorem 6, the set of $p \in \mathbb{P}_k$ for which

$$\operatorname{sgn} S(p) = \operatorname{sgn} T(p)$$

has full density as a subset of $\mathbb{P}_k$. It follows that the set of prime pairs for which $\operatorname{sgn} T(p) = \chi_3(k)$ has lower density

$$\liminf_{x\to\infty} \frac{\{p \in \mathbb{P}_k(x) : \operatorname{sgn} T(p) = \chi_3(k)\}}{\pi_k(x)} \geqslant \prod_{q\in Q} (q-2)^{-1} \left(1 - \frac{R_k}{S_k} - \epsilon\right) + o(1).$$

Because this holds for all $\epsilon > 0$, the lower bound in Theorem 7(a) follows.

5.3. Proof of Theorem 7(b)

As before, we may use $S(p)$ and $T(p)$ interchangeably in what follows. Fix k satisfying $\chi_3(k) = \pm 1$ and let $r \geqslant 5$ be prime. We wish to count the number of $p \in \mathbb{P}_k(x)$ for which $r|(p - 1 + \tau_k)$.

If $r|(k - \chi_3(k))$, then

$$p - 1 + \tau_k \pm (k - \chi_3(k)) \equiv 0 \pmod{r}.$$

Consequently, (5.8) permits us to deduce that $p + k \equiv 0 \pmod{r}$ of $p \equiv 0 \pmod{r}$. In either case, there is at most one such prime p.

Now suppose that $r \nmid (k - \chi_3(k))$ and let

$$g_1(t) = rt + 1 - \tau_k, \quad g_2(t) = rt + 1 + (k - \tau_k), \quad \text{and} \quad g = g_1 g_2.$$

Then

$$N_g(p) = \begin{cases} 0 & \text{if } p = r, \\ 1 & \text{if } p|k, \\ 2 & \text{if } p \nmid k, \end{cases}$$

so the Bateman–Horn conjecture gives

$$\begin{aligned} \sum_{\substack{p\in\mathbb{P}_k(x) \\ p+\tau_k\equiv 1 \pmod{r} \\ r\nmid(k-\chi_3(k))}} 1 &= (1+o(1)) \frac{(x+\tau_k-1)/r}{(\log((x+\tau_k-1)/r))^2} \prod_{p\geqslant 2} \frac{1-N_g(p)/p}{(1-1/p)^2} \\ &= (1+o(1)) \frac{x}{r(\log x)^2} \cdot \frac{1}{(1-1/r)^2} \prod_{p\neq r} \frac{1-N_g(p)/p}{(1-1/p)^2} \\ &= (1+o(1)) \frac{x}{(\log x)^2} \cdot \frac{1}{r-N_f(r)} \prod_{p\geqslant 2} \frac{1-N_f(p)/p}{(1-1/p)^2} \\ &= (1+o(1)) \frac{\pi_k(x)}{r-N_f(r)}, \end{aligned} \tag{5.10}$$

in which $N_f(r)$ refers to (2.2). If $\operatorname{sgn} S(p) = \chi_3(k)$, so that p does not belong to the set of interest in Theorem 7(b), then

$$\frac{1}{2}\prod_{\substack{r|(p-1+\tau_k)\\ r\geqslant 5}}\left(1-\frac{1}{r}\right) = \frac{\varphi(p-1+\tau_k)}{p-1+\tau_k} < \frac{\varphi(p-1+(k-\tau_k))}{p-1+(k-\tau_k)} \leqslant \frac{1}{3},$$

because $3 \nmid (p-1+\tau_k)$ and $3|(p-1+(k-\tau_k))$. If

$$G(p) := \sum_{\substack{r|(p-1+\tau_k)\\ r\geqslant 5}} \log\left(1+\frac{1}{r-1}\right),$$

then $G(p) > \log(3/2)$ whenever $p, p+k$ are primes that satisfy $\operatorname{sgn} S(p) = \chi_3(k)$. Let $\pi_k''(x)$ denote the number of primes $p \in \mathbb{P}_k(x)$ for which $\operatorname{sgn} S(p) > \chi_3(k)$. For sufficiently large x, (5.10) implies that

$$\begin{aligned}
\pi_k''(x)\log(3/2) &< \sum_{p\in\mathbb{P}_k(x)} G(p)\\
&= \sum_{p\in\mathbb{P}_k(x)} \sum_{\substack{r|(p-1+\tau_k)\\ r\geqslant 5}} \log\left(1+\frac{1}{r-1}\right)\\
&\leqslant \sum_{5\leqslant r\leqslant x} \log\left(1+\frac{1}{r-1}\right) \sum_{\substack{p\in\mathbb{P}_k(x)\\ p+\tau_k\equiv 1 \pmod r}} 1\\
&\leqslant \sum_{\substack{5\leqslant r\leqslant x\\ r\nmid(k-\chi_3(k))}} \log\left(1+\frac{1}{r-1}\right) \sum_{\substack{p\in\mathbb{P}_k(x)\\ p+k\equiv 1 \pmod r}} 1\\
&\qquad + \sum_{\substack{5\leqslant r\leqslant x\\ r|(k-\chi_3(k))}} \log\left(1+\frac{1}{r-1}\right)\\
&\leqslant \Bigg[(1+o(1)) \sum_{\substack{r\geqslant 5\\ r\nmid(k-\chi_3(k))}} \frac{1}{r-N_f(r)} \log\left(1+\frac{1}{r-1}\right)\\
&\qquad + \frac{1}{\pi_k(x)} \sum_{\substack{r\geqslant 5\\ r|(k-\chi_3(k))}} \log\left(1+\frac{1}{r-1}\right)\Bigg] \pi_k(x)\\
&= \left(\sum_{\substack{r\geqslant 5\\ r\nmid(k-\chi_3(k))}} \frac{1}{r-N_f(r)} \log\left(1+\frac{1}{r-1}\right) + o(1)\right) \pi_k(x)\\
&= (R_k' + o(1))\,\pi_k(x).
\end{aligned}$$

Thus, there are at least

$$\pi_k(x) - \pi_k''(x) > \pi_k(x)\left(1 - \frac{R_k'}{\log(3/2)} - o(1)\right)$$

primes $p \in \mathbb{P}_k(x)$ such $\operatorname{sgn} S(p) = -\chi_3(k)$. Reasoning similar to that used in the conclusion of the proof of part (a) yield the formula in Theorem 7(b).

To show that this lower density is bounded below by 0.6515, we observe that[2]

$$\begin{aligned} R_k' &= \sum_{\substack{r \geqslant 5 \\ r \nmid (k - \chi_3(k))}} \frac{1}{r - N_f(r)} \log\left(1 + \frac{1}{r-1}\right) \\ &\leqslant \sum_{r \geqslant 5} \frac{1}{r-2} \log\left(1 + \frac{1}{r-1}\right) \\ &< 0.1412981. \end{aligned}$$

It follows that

$$1 - \frac{R_k'}{\log(3/2)} > 1 - \frac{0.1412981}{\log(3/2)} > 0.6515.$$

This completes the proof of Theorem 7. □

6. Extending Theorem 7 to pairs $p, p+k$ with $k \equiv 0 \pmod 3$

Fix $k \equiv 0 \pmod 3$. The techniques used in the proof of Theorem 7(a) can be used to show that $T(p) < 0$ and $T(p) > 0$ both occur with positive density as a subset of $\mathbb{P}_k$. Because the proofs are nearly identical, we simply point out the small differences and leave the remaining details to the reader.

Since $\chi_3(k) = 0$ whenever $k \equiv 0 \pmod 3$, some notational adjustment is needed. To show that $T(p) < 0$ occurs with positive density in $\mathbb{P}_k$, we follow the proof of Theorem 7(a) as if $k \equiv -1 \pmod 3$, replacing each occurrence of $\chi_3(k)$ with -1. Similarly, to show that $T(p) > 0$ occurs with positive density, we follow the proof as if $k \equiv 1 \pmod 3$, replacing $\chi_3(k)$ with 1.

We modify the definition of L_k by setting

$$L_k^{\pm} := \log\left[\prod_{q \in Q^{\pm}} \left(1 + \frac{1}{q-1}\right)\right],$$

in which $Q^{\pm}$ are finite sets of primes to be determined shortly. Note the absence of the 2/3 factor inside the logarithm. This is due to the fact that 3 either divides both $p-1$ and

[2] The terms of R_k' are $O(1/r^2)$, since $\log(1+t) < t$ for $t > 0$, so the series converges. Mathematica provides the numerical value 0.141298112.

Table 5
Lower bounds from Theorem 10 for $k \equiv 0 \pmod 3$.

k	$\chi_3(k) = -1$				$\chi_3(k) = 1$			
	Q^-	L_k^-	R_k^-	Lwr bd: $T(p) < 0$	Q^+	L_k^+	R_k^+	Lwr bd: $T(p) > 0$
6	5	0.223144	0.066917	0.233372	7	0.154151	0.110468	0.056675
12	5	0.223144	0.056327	0.249192	5	0.223144	0.059640	0.244242
18	5	0.223144	0.062875	0.23941	5	0.223144	0.063737	0.238123
24	7	0.154151	0.108351	0.059422	5	0.223144	0.066917	0.233372
30	7	0.154151	0.090573	0.082487	7	0.154151	0.090742	0.082268
36	5	0.223144	0.036087	0.279427	11, 13	0.175353	0.122649	0.003035
42	5	0.223144	0.061145	0.241994	5	0.223144	0.061205	0.241905
48	5	0.223144	0.066439	0.234086	5	0.223144	0.036087	0.279427
54	7	0.154151	0.110094	0.057159	5	0.223144	0.056327	0.249192
60	7	0.154151	0.091573	0.081190	7	0.154151	0.091593	0.081164
66	5	0.223144	0.058581	0.245824	7	0.154151	0.109178	0.058349
72	5	0.223144	0.066711	0.233679	5	0.223144	0.066723	0.233663
78	5	0.223144	0.024890	0.296152	5	0.223144	0.066145	0.234525
84	11, 13	0.175353	0.118143	0.003295	5	0.223144	0.057737	0.247086
90	11, 17	0.155935	0.107941	0.002279	7	0.154151	0.084596	0.090242
96	5	0.223144	0.063737	0.238123	7	0.154151	0.110359	0.056816
102	5	0.223144	0.066564	0.2339	5	0.223144	0.066568	0.233894
108	5	0.223144	0.066828	0.233506	5	0.223144	0.066831	0.233501
114	7	0.154151	0.110211	0.057008	5	0.223144	0.064624	0.236798
120	7	0.154151	0.087831	0.086045	11, 13	0.175353	0.104836	0.004062
126	5	0.223144	0.061779	0.241048	11, 13	0.175353	0.11823	0.003290
132	5	0.223144	0.065799	0.235043	5	0.223144	0.031847	0.28576
138	5	0.223144	0.066766	0.233597	5	0.223144	0.066768	0.233595
144	7	0.154151	0.092601	0.079856	5	0.223144	0.065617	0.235314
150	7	0.154151	0.091827	0.080860	7	0.154151	0.091828	0.080859
156	5	0.223144	0.065180	0.235967	7	0.154151	0.10982	0.0575156

$p+k-1$, or it divides neither. Consequently, the usual 2/3 from (5.3) is "canceled" when we compare $\varphi(p-1)/(p-1)$ and $\varphi(p+k-1)/(p+k-1)$. This is also the reason why we cannot employ the techniques from the proof Theorem 7(b) to establish a lower density greater than 0.5 when $k \equiv 0 \pmod 3$. This is not surprising, since Table 1 demonstrates that there is no universal bias in the sign of $T(p)$ that applies for all $k \equiv 0 \pmod 3$.

Next, we let

$$R_k^{\pm} = \sum_{\substack{r \geqslant 5 \\ r \notin Q^{\pm} \\ r \nmid (k \pm 1)}} \frac{1}{r - N_f(r)} \log\left(1 + \frac{1}{r-1}\right),$$

in which the signs are chosen depending on whether we wish to prove $T(p) > 0$ or $T(p) < 0$. We define $Q^{\pm}$ to be the smallest ordered subset of primes for which $q \nmid k(k \mp 1)$ for all $q \in Q^{\pm}$ and such that

$$L_k^{\pm} > R_k^{\pm}.$$

Beyond the aforementioned, the only other difference in the proof is the absence of the 2/3 factor when comparing $\varphi(p-1)/(p-1)$ and $\varphi(p+k-1)/(p+k-1)$. With this in mind, we have the following result.

Theorem 10. *Assume that the Bateman–Horn conjecture holds. If* $k \equiv 0 \pmod 3$*, then the set of primes* $p \in \mathbb{P}_k$ *for which*

$$\operatorname{sgn} T(p) = \pm 1$$

has lower density (as a subset of $\mathbb{P}_k$*) at least*

$$\prod_{q \in Q^{\pm}} (q-2)^{-1} \left(1 - \frac{R_k^{\pm}}{L_k^{\pm}}\right) > 0.$$

Table 5 provides numerical values for $R_k^{\pm}$, $L_k^{\pm}$, and the bounds in Theorem 10 for various values of $k \equiv 0 \pmod 3$.

References

[1] Paul T. Bateman, Roger A. Horn, A heuristic asymptotic formula concerning the distribution of prime numbers, Math. Comp. 16 (1962) 363–367.

[2] Wouter Castryck, Étienne Fouvry, Gergely Harcos, Emmanuel Kowalski, Philippe Michel, Paul Nelson, Eytan Paldi, János Pintz, Andrew V. Sutherland, Terence Tao, Xiao-Feng Xie, New equidistribution estimates of Zhang type, Algebra Number Theory 8 (9) (2014) 2067–2199.

[3] S.R. Garcia, E. Kahoro, F. Luca, Primitive root bias for twin primes, Exp. Math. (2017), https://doi.org/10.1080/10586458.2017.1360809, in press, http://arxiv.org/abs/1705.02485.

[4] S.W. Graham, Jeffrey J. Holt, Carl Pomerance, On the solutions to $\varphi(n) = \varphi(n+k)$, in: Number Theory in Progress, vol. 2, Zakopane–Kościelisko, 1997, de Gruyter, Berlin, 1999, pp. 867–882.

[5] James Maynard, Small gaps between primes, Ann. of Math. (2) 181 (1) (2015) 383–413.

[6] József Sándor, Dragoslav S. Mitrinović, Borislav Crstici, Handbook of Number Theory. I, Springer, Dordrecht, 2006, second printing of the 1996 original.

[7] Gérald Tenenbaum, Introduction to Analytic and Probabilistic Number Theory, third ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015, translated from the 2008 French edition by Patrick D.F. Ion.

[8] Axel Thue, Selected Mathematical Papers, Universitetsforlaget, Oslo, 1977, with an introduction by Carl Ludwig Siegel and a biography by Viggo Brun, edited by Trygve Nagell, Atle Selberg, Sigmund Selberg, and Knut Thalberg.

[9] R. Tijdeman, On integers with many small prime factors, Compos. Math. 26 (1973) 319–330.

[10] Tomohiro Yamada, On equations $\sigma(n) = \sigma(n+k)$ and $\varphi(n) = \varphi(n+k)$, J. Comb. Number Theory (2017), in press, https://arxiv.org/1001.2511.

[11] Yitang Zhang, Bounded gaps between primes, Ann. of Math. (2) 179 (3) (2014) 1121–1174.