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

Stephan Ramon Garcia\textsuperscript{a,\ast}, Florian Luca\textsuperscript{b,c,d}, Timothy Schaaff\textsuperscript{a}

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

\begin{abstract}
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.

\textcopyright{} 2017 Elsevier Inc. All rights reserved.
\end{abstract}
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 $p \leq x$ 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 $\pi_2$ 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 $\pi_k$ 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 $k > 4$. 
(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 $k \geq 4$ 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 $k \geq 4$ 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 $T(p) \leq 0$, a proportion of $1.76 \times 10^{-6}$. 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 $1.81 \times 10^{-20}$) 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).

<table>
<thead>
<tr>
<th>$k$</th>
<th>$#T(p) &lt; 0$</th>
<th>$\pi_k(x)$ Proportion</th>
<th>$#T(p) &lt; 0$</th>
<th>$\pi_k(x)$ Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>28490</td>
<td>1418478 0.0200849</td>
<td>62</td>
<td>1980 0.9819111</td>
</tr>
<tr>
<td>4</td>
<td>1390701</td>
<td>1419044 0.980027</td>
<td>64</td>
<td>1416847 1418937 0.998527</td>
</tr>
<tr>
<td>6</td>
<td>1682707</td>
<td>2836640 0.594791</td>
<td>66</td>
<td>2187908 3153911 0.693713</td>
</tr>
<tr>
<td>8</td>
<td>28771</td>
<td>1417738 0.0202903</td>
<td>68</td>
<td>25409 1512639 0.0167978</td>
</tr>
<tr>
<td>10</td>
<td>1891800</td>
<td>1891902 0.999946</td>
<td>70</td>
<td>2270424 2270424 1</td>
</tr>
<tr>
<td>12</td>
<td>1441259</td>
<td>2837946 0.507853</td>
<td>72</td>
<td>1431789 2837290 0.504649</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1703216 1.76 $\times$ $10^{-6}$</td>
<td>74</td>
<td>64 1459313 4.39 $\times$ $10^{-5}$</td>
</tr>
<tr>
<td>16</td>
<td>1420209</td>
<td>1420273 0.999955</td>
<td>76</td>
<td>1503210 1502338 0.999981</td>
</tr>
<tr>
<td>18</td>
<td>1433488</td>
<td>2837906 0.505122</td>
<td>78</td>
<td>1745211 3096187 0.563665</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>1891034 2.12 $\times$ $10^{-6}$</td>
<td>80</td>
<td>113 1829585 5.9 $\times$ $10^{-5}$</td>
</tr>
<tr>
<td>22</td>
<td>1576076</td>
<td>1576379 0.999980</td>
<td>82</td>
<td>1426536 1455721 0.979952</td>
</tr>
<tr>
<td>24</td>
<td>1015032</td>
<td>2838360 0.357612</td>
<td>84</td>
<td>1145652 3404217 0.336539</td>
</tr>
<tr>
<td>26</td>
<td>26521</td>
<td>1546675 0.0171471</td>
<td>86</td>
<td>1547184 1572758 0.982635</td>
</tr>
<tr>
<td>28</td>
<td>1495332</td>
<td>1702838 0.998206</td>
<td>88</td>
<td>1530480 3784105 0.510155</td>
</tr>
<tr>
<td>30</td>
<td>1930480</td>
<td>3784105 0.510155</td>
<td>90</td>
<td>1489160 3785003 0.393437</td>
</tr>
<tr>
<td>32</td>
<td>20553</td>
<td>1418579 0.0144884</td>
<td>92</td>
<td>29413 1486659 0.0197846</td>
</tr>
<tr>
<td>34</td>
<td>1495332</td>
<td>1513933 0.987713</td>
<td>94</td>
<td>1421558 1450180 0.980263</td>
</tr>
<tr>
<td>36</td>
<td>2097416</td>
<td>2838465 0.738926</td>
<td>96</td>
<td>1575169 2839516 0.674682</td>
</tr>
<tr>
<td>38</td>
<td>21739</td>
<td>1502517 0.0144684</td>
<td>98</td>
<td>377 1702580 0.000221429</td>
</tr>
<tr>
<td>40</td>
<td>1891651</td>
<td>1891659 0.999996</td>
<td>100</td>
<td>1891334 1891337 0.999998</td>
</tr>
<tr>
<td>42</td>
<td>1727098</td>
<td>3405081 0.507212</td>
<td>102</td>
<td>1531067 3027395 0.505737</td>
</tr>
<tr>
<td>44</td>
<td>6</td>
<td>1576157 3.81 $\times$ $10^{-6}$</td>
<td>104</td>
<td>9 1549054 5.8 $\times$ $10^{-6}$</td>
</tr>
<tr>
<td>46</td>
<td>1486910</td>
<td>1486946 0.999976</td>
<td>106</td>
<td>1447486 1447486 1</td>
</tr>
<tr>
<td>48</td>
<td>1318068</td>
<td>2838746 0.464313</td>
<td>108</td>
<td>1434316 2838777 0.505258</td>
</tr>
<tr>
<td>50</td>
<td>48</td>
<td>1891847 2.54 $\times$ $10^{-5}$</td>
<td>110</td>
<td>16 2101919 7.6 $\times$ $10^{-6}$</td>
</tr>
<tr>
<td>52</td>
<td>1525943</td>
<td>1548356 0.985525</td>
<td>112</td>
<td>1699877 1702796 0.998266</td>
</tr>
<tr>
<td>54</td>
<td>933772</td>
<td>2839928 0.328801</td>
<td>114</td>
<td>1051285 3005470 0.349895</td>
</tr>
<tr>
<td>56</td>
<td>2272</td>
<td>1701628 0.00133519</td>
<td>116</td>
<td>22762 1417107 0.0154736</td>
</tr>
<tr>
<td>58</td>
<td>1447184</td>
<td>1472758 0.982635</td>
<td>118</td>
<td>1418455 1442208 0.98353</td>
</tr>
<tr>
<td>60</td>
<td>1939665</td>
<td>3783957 0.512602</td>
<td>120</td>
<td>2269102 3784749 0.599538</td>
</tr>
</tbody>
</table>

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 $C_k$.

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

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

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

For a fixed $k$, let

$$f(t) = t(t + k),$$

so that

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

The Bateman–Horn conjecture predicts that

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

in which

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

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

$$P(x) \leq \frac{BC}{D} \int_{2}^{x} \frac{dt}{(\log t)^m} = (1 + o(1)) \frac{BC}{D} \frac{x}{(\log x)^m}$$

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

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

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

3. Totient comparison theorem

The well-known formula

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

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 $p \geq 5$ [3, Lem. 2]. For $k \geq 4$, 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

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

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 \( n \leq x \) 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 \( n_i \geq 3 \). 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 \( \pi_k \).

**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

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

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 \leq \frac{g(x-j+1)}{j(j+k)} \tag{so that \( p \leq x \)}
\]
for which
\[ p = j \left( \frac{j + k}{g} t + 1 \right) + 1, \quad q = j \left( \frac{j + k}{g} t + 1 \right) + k + 1, \]
\[ r = \frac{j}{g} t = 1, \quad s = \frac{j + k}{g} t + 1, \]
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 \( o(\pi_k(x)) \). \( \square \)

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 \( p - 1 \leq x \) and \( P(p - 1) \leq k \), then \( p - 1 \) is divisible only by the \( \pi(k) \) primes at most \( k \). The number of such \( p \) at most \( x \) is \( O((\log x)^{\pi(k)}) = o(x/(\log x)^2) \), 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 \( \ell \geq 1 \), 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.** \( \sum_{q^a \leq x} \frac{1}{q^a} = \log \log x + O(1) \).

**Proof.** Mertens’ theorem [6, §VII.28.1b] implies that
\[ \sum_{q \leq x} \frac{1}{q} = \log \log x + O(1). \]
Thus,
\[ \sum_{q^a \leq x} \frac{1}{q^a} = \sum_{q \leq x} \frac{1}{q} + \sum_{q^a \leq x} \frac{1}{q^a} \leq \log \log x + O(1) + \sum_{n \geq 2} \sum_{k \geq 2} \frac{1}{n^k} \]
\[ = \log \log x + O(1) + \sum_{n \geq 2} \frac{1}{n^2} \cdot \frac{1}{1 - 1/n} \]
\[ = \log \log x + O(1). \] \( \square \)

1 If \( \pi(k) = s \) and \( P(p - 1) \leq k \), 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 \( p_i^{a_i} \leq x \) and hence \( a_i \leq (\log x)/\log p_i \). 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

\[
\varphi(n) = \prod_{p^a \mid n} p^{a-1}(p - 1)
\]

ensures that \( 2^{\omega(n) - 1} \mid \varphi(n) \) 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 \( k \geq 2 \) even and \( \ell \geq 1 \),

\[
\# \{ p \in \mathbb{P}_k(x) : 2^\ell \mid T(p) \} \sim \pi_k(x).
\]

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

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

\[
g_1(t) = t, \quad g_2(t) = nt + 1, \quad g_3(t) = nt + k + 1.
\]

Then

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

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

\[
\sum_{\substack{t \leq \frac{x}{n} \sum_{\substack{t,n \leq \frac{x}{n} \text{ prime}}}} 1 \leq \frac{x}{n \log x} \prod_{q} \frac{1 - N_g(q)/q}{1/q} \\
\leq \frac{x}{n \log x} \prod_{q \mid n} \frac{1 - 1/q}{1 - 1/q} \prod_{q \mid k \text{ or } q \mid (k + 1)} \frac{1 - 2/q}{1 - 1/q} \prod_{q \mid n \text{ and } q \mid k} \frac{1 - 3/q}{1 - 1/q} \\
\leq \frac{x}{n \log x} \left[ \frac{1}{(1 - 1/2)^2} \right]^{\omega(n)} \prod_{q \mid k \text{ or } q \mid (k + 1)} \frac{1 - 2/q}{1 - 1/q} \prod_{q \mid n \text{ and } q \mid k} \frac{1 - 3/q}{1 - 1/q} \quad (3.4)
\]

\[
\leq \frac{2^{2\ell} x}{n \log x}.
\]
In the preceding computation, we used the fact that
\[
1 \leq \frac{1 - 2/q}{(1 - 1/q)^3} \quad \text{for } q \geq 3
\]
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
\[
\sum_{p \in \mathbb{P}_k(x)} \sum_{\omega(p - 1) \leq \ell} 1 = \sum_{n \leq x} \sum_{\omega(n) \leq \ell} \sum_{t, n+1 \text{ prime}} 1 \\
\ll \sum_{n \leq x} \frac{2^{2\ell} x}{n (\log x)^3} \\
\ll \frac{x}{(\log x)^3} \sum_{n \leq x} \frac{1}{n} \\
\ll \frac{x}{(\log x)^3} \left( 1 + \sum_{q^a \leq x} \frac{1}{q^a} \right)^\ell \\
\ll \frac{x}{(\log x)^3} \left( \log \log x + O(1) \right)^\ell \\
\ll o(\pi_k(x)).
\]
Similarly, the count of \(p \in \mathbb{P}_k(x)\) with \(\omega(p + k - 1) \leq \ell\) is also \(o(\pi_k(x))\). \(\square\)

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

$$0 < (p + k - 1)\phi(p - 1) - (p - 1)\phi(p + k - 1)$$

$$= p(\phi(p - 1) - \phi(p + k - 1)) + (k - 1)\phi(p - 1) + \phi(p + k - 1)$$

$$\leq p(\phi(p - 1) - \phi(p + k - 1)) + (k - 1)(p - 1) + (p + k - 1)$$

$$\leq p(\phi(p - 1) - \phi(p + k - 1) + k)$$

$$= p(T(p) + k). \quad (3.5)$$

Fix $\ell$ so that $2^{\ell} \geq k$. Apply Lemma 5 at (3.5) and conclude that

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

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

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

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

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 $p \geq 5$. Then (3.6) becomes

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

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
\[
\frac{\varphi(p - 1)}{p - 1} = \frac{1}{3} \prod_{q \mid (p - 1), q \geq 5} \left( 1 - \frac{1}{q} \right)
\quad \text{and} \quad
\frac{\varphi(p + k - 1)}{p + k - 1} = \frac{1}{2} \prod_{q \mid (p + k - 1), q \mid (k+1)} \left( 1 - \frac{1}{q} \right).
\]
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 \( k \equiv 0 \pmod{3} \), 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 \( q \geq 5 \) 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 \( 3 \mid k \).

5. Primitive roots biases for \( k \not\equiv 0 \pmod{3} \)

Let \( \chi_3 \) 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
\[
5 \leq q_1 < q_2 < q_3 < \cdots
\]
be the ordered sequence of primes that do not divide
\[
k(k - \chi_3(k)),
\]
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
\[
Q = \{q_1, q_2, \ldots, q_m\},
\]
in which the index \( m \) shall be determined momentarily. Define
\[
L_k = \log \left[ \frac{2}{3} \prod_{q \in Q} \left( 1 + \frac{1}{q - 1} \right) \right]
\]
and
\[
R_k = \sum_{\substack{r \geq 5 \\ r \in Q \setminus \{r \mid (k, r) = 1\}}} \frac{1}{r - N_f(r)} \log \left( 1 + \frac{1}{r - 1} \right),
\]

(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

\[
L_k > R_k.
\]

(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

\[
R'_k = \sum_{\substack{r \geq 5 \\ r \mid (k, r - 3) \mid k}} \frac{1}{r - N_f(r)} \log \left( 1 + \frac{1}{r - 1} \right).
\]

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

\[
\text{sgn} T(p) = \chi_3(k)
\]

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

\[
\text{sgn} T(p) = -\chi_3(k)
\]

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

\[
1 - \frac{R'_k}{\log(3/2)} > 0.6515.
\]
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 $k \equiv 1 \pmod{3}$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$Q$</th>
<th>$L_k$</th>
<th>$R_k$</th>
<th>Lower bound: $T(p) &gt; 0$</th>
<th>$R_k^*$</th>
<th>Lower bound: $T(p) &lt; 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.025497</td>
<td>0.004594</td>
<td>0.141298</td>
<td>0.651516</td>
</tr>
<tr>
<td>10</td>
<td>7, 11, 13, 17, 19, 23</td>
<td>0.083221</td>
<td>0.066667</td>
<td>$8.39 \times 10^{-8}$</td>
<td>0.122703</td>
<td>0.697378</td>
</tr>
<tr>
<td>16</td>
<td>7, 11, 13, 17, 19, 23, 29</td>
<td>0.118273</td>
<td>0.081693</td>
<td>$4.28 \times 10^{-9}$</td>
<td>0.066917</td>
<td>0.834963</td>
</tr>
<tr>
<td>22</td>
<td>5, 13, 17, 19, 23</td>
<td>0.056865</td>
<td>0.049234</td>
<td>$7.58 \times 10^{-7}$</td>
<td>0.109409</td>
<td>0.730164</td>
</tr>
<tr>
<td>28</td>
<td>5, 11, 13, 17</td>
<td>0.053656</td>
<td>0.038571</td>
<td>0.000603</td>
<td>0.13616</td>
<td>0.664189</td>
</tr>
<tr>
<td>34</td>
<td>5, 7, 13</td>
<td>0.051872</td>
<td>0.028558</td>
<td>0.002723</td>
<td>0.130455</td>
<td>0.678257</td>
</tr>
<tr>
<td>40</td>
<td>7, 11, 13, 17, 19, 23, 29, 31</td>
<td>0.071021</td>
<td>0.068880</td>
<td>$1.59 \times 10^{-10}$</td>
<td>0.115426</td>
<td>0.715324</td>
</tr>
<tr>
<td>46</td>
<td>7, 11, 13, 17, 19, 23, 29, 31</td>
<td>0.106611</td>
<td>0.082375</td>
<td>$2.30 \times 10^{-9}$</td>
<td>0.066821</td>
<td>0.8352</td>
</tr>
<tr>
<td>52</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.024517</td>
<td>0.004702</td>
<td>0.13665</td>
<td>0.662979</td>
</tr>
<tr>
<td>58</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.025150</td>
<td>0.004632</td>
<td>0.138071</td>
<td>0.659474</td>
</tr>
<tr>
<td>64</td>
<td>5, 11, 13, 17</td>
<td>0.053656</td>
<td>0.045009</td>
<td>0.000036</td>
<td>0.110468</td>
<td>0.727552</td>
</tr>
<tr>
<td>70</td>
<td>11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61</td>
<td>0.102261</td>
<td>0.084619</td>
<td>$1.81 \times 10^{-20}$</td>
<td>0.115448</td>
<td>0.715271</td>
</tr>
<tr>
<td>76</td>
<td>7, 11, 13, 17, 19, 23, 29, 31</td>
<td>0.096996</td>
<td>0.083859</td>
<td>$1.11 \times 10^{-9}$</td>
<td>0.066740</td>
<td>0.835398</td>
</tr>
<tr>
<td>82</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.025331</td>
<td>0.004612</td>
<td>0.141282</td>
<td>0.651555</td>
</tr>
<tr>
<td>88</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.027621</td>
<td>0.002833</td>
<td>0.138937</td>
<td>0.657333</td>
</tr>
<tr>
<td>94</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.022306</td>
<td>0.004946</td>
<td>0.140157</td>
<td>0.65433</td>
</tr>
<tr>
<td>100</td>
<td>7, 13, 17, 19, 23, 29, 31, 37</td>
<td>0.083152</td>
<td>0.071944</td>
<td>$1.67 \times 10^{-11}$</td>
<td>0.112113</td>
<td>0.723496</td>
</tr>
<tr>
<td>106</td>
<td>11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 59</td>
<td>0.111135</td>
<td>0.108798</td>
<td>$3.52 \times 10^{-19}$</td>
<td>0.036080</td>
<td>0.91017</td>
</tr>
<tr>
<td>112</td>
<td>5, 11, 13, 17</td>
<td>0.053656</td>
<td>0.039790</td>
<td>0.000058</td>
<td>0.135377</td>
<td>0.66119</td>
</tr>
<tr>
<td>118</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.02145</td>
<td>0.005040</td>
<td>0.134016</td>
<td>0.669475</td>
</tr>
<tr>
<td>124</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.025459</td>
<td>0.004599</td>
<td>0.140627</td>
<td>0.65317</td>
</tr>
<tr>
<td>130</td>
<td>7, 11, 17, 19, 23, 29, 31</td>
<td>0.071021</td>
<td>0.068848</td>
<td>$1.62 \times 10^{-10}$</td>
<td>0.121523</td>
<td>0.700289</td>
</tr>
<tr>
<td>136</td>
<td>7, 11, 13, 17, 19, 23, 29, 31</td>
<td>0.099439</td>
<td>0.084567</td>
<td>$4.69 \times 10^{-10}$</td>
<td>0.066904</td>
<td>0.835586</td>
</tr>
<tr>
<td>142</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.018217</td>
<td>0.005397</td>
<td>0.140817</td>
<td>0.652702</td>
</tr>
<tr>
<td>148</td>
<td>5, 11, 13, 17</td>
<td>0.053656</td>
<td>0.044941</td>
<td>0.000036</td>
<td>0.110464</td>
<td>0.727606</td>
</tr>
<tr>
<td>154</td>
<td>5, 13, 19, 23, 29, 31</td>
<td>0.064121</td>
<td>0.045715</td>
<td>$3.11 \times 10^{-8}$</td>
<td>0.131059</td>
<td>0.676768</td>
</tr>
<tr>
<td>160</td>
<td>7, 11, 13, 17, 19, 23</td>
<td>0.083182</td>
<td>0.064667</td>
<td>$8.39 \times 10^{-8}$</td>
<td>0.122329</td>
<td>0.698299</td>
</tr>
<tr>
<td>166</td>
<td>7, 13, 17, 19, 23, 29, 31, 37, 41</td>
<td>0.107845</td>
<td>0.089968</td>
<td>$5.26 \times 10^{-13}$</td>
<td>0.056325</td>
<td>0.861085</td>
</tr>
<tr>
<td>172</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.025449</td>
<td>0.004599</td>
<td>0.138104</td>
<td>0.659394</td>
</tr>
<tr>
<td>178</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.025464</td>
<td>0.004588</td>
<td>0.140997</td>
<td>0.652259</td>
</tr>
<tr>
<td>184</td>
<td>5, 7, 11</td>
<td>0.067139</td>
<td>0.024618</td>
<td>0.004691</td>
<td>0.140922</td>
<td>0.652444</td>
</tr>
</tbody>
</table>

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

$$q \nmid k(k + 1) \quad \text{(resp., } q \nmid k(k - 1)),$$
for all \( q \in \mathbb{Q} \). The number of \( p \in \mathbb{P}_k(x) \) such that
\[
q | (p - 1) \quad \text{(resp., } q | (p + k - 1))
\]
for all \( q \in \mathbb{Q} \) is
\[
\pi_k'(x) = (1 + o(1)) \pi_k(x) \prod_{q \in \mathbb{Q}} (q - 2)^{-1}.
\]

**Proof.** Suppose that \( q \nmid k(k+1) \) for all \( q \in \mathbb{Q} \), 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 \( q \in \mathbb{Q} \). If \( a = \prod_{q \in \mathbb{Q}} q \), then the desired primes are those of the form
\[
n = at + 1 \leq x \quad \text{such that} \quad n + k = at + k + 1 \text{ is prime}.
\]
Let
\[
g_1(t) = at + 1, \quad g_2(t) = at + k + 1, \quad \text{and} \quad g = g_1 g_2.
\]
In the Bateman–Horn conjecture with \( s \) denoting an arbitrary prime, we have
\[
N_g(s) = \begin{cases} 
0 & \text{if } s \in \mathbb{Q}, \\
1 & \text{if } s | k, \ s \notin \mathbb{Q}, \\
2 & \text{if } s \nmid k, \ s \notin \mathbb{Q}.
\end{cases}
\]
(5.6)

For sufficiently large \( x \), the Bateman–Horn conjecture predicts that the number of such \( t \leq (x - 1)/a \) is
\[
\pi_k'(x) = (1 + o(1)) \frac{(x - 1)/a}{(\log((x - 1)/a))^2} \prod_{s \geq 2} \frac{1 - N_g(s)/s}{(1 - 1/s)^2}
\]
\[
= (1 + o(1)) \frac{x}{a(\log x)^2} \prod_{s \geq 2} \frac{1 - N_g(s)/s}{(1 - 1/s)^2}
\]
\[
= (1 + o(1)) \frac{x}{a(\log x)^2} \prod_{s \in \mathbb{Q}} \frac{1}{(1 - 1/s)^2} \prod_{s \notin \mathbb{Q}} \frac{1 - N_g(s)/s}{(1 - 1/s)^2}
\]
\[
= (1 + o(1)) \frac{x}{a(\log x)^2} \prod_{s \in \mathbb{Q}} (1 - N_f(s)/s)^{-1} \prod_{s \geq 2} \frac{1 - N_f(s)/s}{(1 - 1/s)^2},
\]
in which \( N_f(s) \) refers to (2.2). Simplifying further yields
\[ \pi'_k(x) = (1 + o(1)) \frac{\pi_k(x)}{a} \prod_{s \in Q} (1 - 2/s)^{-1} \]

\[ = (1 + o(1))\pi_k(x) \prod_{q \in Q} (q - 2)^{-1}. \quad \square \]

**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

\[ q \nmid k(k + 1) \quad (\text{resp., } q \nmid k(k - 1)), \]

for all \( q \in Q \). Let \( r \geq 5 \) be a fixed prime not in \( Q \) that satisfies

\[ r \nmid k(k - 1) \quad (\text{resp., } r \nmid k(k + 1)). \]

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

\[ q|(p - 1) \quad \text{and} \quad r|(p + k - 1) \quad (\text{resp., } r|(p - 1)), \]

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 \( r \geq 5 \) such that \( r \nmid (k - 1) \) and let \( a = \prod_{q \in Q} q \). The desired primes are precisely those of the form

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

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

\[ n = ar\ell + b_r \quad \text{and} \quad n + k = ar\ell + (b_r + k) \quad (5.7) \]

are both prime, \( n \leq x \), and

\[ \ell \leq \frac{x - b_r}{ar}. \]
In the Bateman–Horn conjecture, let
\[ g_1(t) = art + b_r, \quad g_2(t) = art + (b_r + k), \quad \text{and} \quad g = g_1g_2. \]

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,
\[ g_1(t) \equiv b_r \equiv aj_0 + 1 \equiv -k + 1 \not\equiv 0 \quad (\text{mod } r) \quad \text{and} \quad g_2(t) \equiv b_r + k \equiv 1 \quad (\text{mod } r) \]
for all \( t \). As \( x \to \infty \), the Bateman–Horn conjecture predicts that the number of such \( \ell \) is
\[
\pi_{k,r}(x) = (1 + o(1)) \frac{(x - b_r)/(ar)}{\log((x - b_r)/(ar))^2} \prod_{s \geq 2} \left( \frac{1 - N_g(s)/s}{(1 - 1/s)^2} \right) \\
= (1 + o(1)) \frac{x}{ar(\log x)^2} \prod_{s \geq 2} \left( \frac{1 - N_g(s)/s}{(1 - 1/s)^2} \right) \\
= (1 + o(1)) \frac{x}{ar(\log x)^2} \prod_{s \in Q \text{ or } s \neq r} \left( \frac{1}{(1 - 1/s)^2} \right) \prod_{s \in Q, s \neq r} \left( \frac{1 - N_g(s)/s}{(1 - 1/s)^2} \right) \\
= (1 + o(1)) \frac{x}{ar(\log x)^2} \prod_{s \in Q \text{ or } s \neq r} (1 - N_f(s)/s)^{-1} \prod_{s \geq 2} \left( \frac{1 - N_f(s)/s}{(1 - 1/s)^2} \right) \\
= (1 + o(1)) \frac{\pi_k(x)}{r - N_f(r)} \prod_{q \in Q} (q - 2)^{-1}.
\]

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 \( k \not\equiv 0 \quad (\text{mod } 3) \).

- 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,
\[
\tau_k = \begin{cases} 
0 & \text{if } \chi_3(k) = -1, \\
k & \text{if } \chi_3(k) = 1,
\end{cases} \quad (5.8)
\]
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 $r \geq 5$ 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 $q \nmid (p - 1 + \tau_k)$ for all $q \in Q$. Thus,

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

so that

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

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

$$\frac{1}{3} \prod_{\substack{r|(p - 1 + (k - \tau_k)) \\ r \geq 5, r \not\in Q}} \left(1 - \frac{1}{r}\right) = \frac{\varphi(p - 1 + (k - \tau_k))}{p - 1 + (k - \tau_k)}$$

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

Consequently,

$$\prod_{\substack{r|(p - 1 + (k - \tau_k)) \\ r \geq 5, r \not\in Q}} \left(1 + \frac{1}{r - 1}\right) > \frac{2}{3} \prod_{q \in Q} \left(1 + \frac{1}{q - 1}\right),$$

in which $r$ is prime. Let

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

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

\[
A(x) = \sum_{p \text{ counted by } \pi_k(x)} F(p)
\]

\[
= \sum_{r \geq 5, r \notin Q} \pi'_{k,r}(x) \log \left( 1 + \frac{1}{r - 1} \right)
\]

\[
\leq \sum_{5 \leq r \leq z, r \notin Q} \pi'_{k,r}(x) \log \left( 1 + \frac{1}{r - 1} \right)
\]

\[
+ \sum_{z < r \leq (\log x)^3} \pi'_{k,r}(x) \log \left( 1 + \frac{1}{r - 1} \right)
\]

\[
+ \sum_{(\log x)^3 < r \leq x} \pi'_{k,r}(x) \log \left( 1 + \frac{1}{r - 1} \right)
\]

\[
= A_1(x) + A_2(x) + A_3(x),
\]  

(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 \( 5 \leq r \leq z \) 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 \( \pi'_{k,r}(x) \leq 1 \) when \( r | (k + \chi_3(k)) \). Thus, for sufficiently large \( x \) we have
\[
A_1(x) \leq (1 + o(1)) \frac{\pi_k(x)}{\prod_{q \in Q} (q - 2)} \sum_{5 \leq r \leq z} \frac{1}{r - N_f(r)} \log \left(1 + \frac{1}{r - 1}\right) \\
+ \sum_{5 \leq r \leq z} \frac{1}{r - N_f(r)} \log \left(1 + \frac{1}{r - 1}\right) \\
\leq (1 + o(1)) \frac{R_k \pi_k(x)}{\prod_{q \in Q} (q - 2)} + \sum_{5 \leq r \leq z} \frac{1}{r - N_f(r)} \log \left(1 + \frac{1}{r - 1}\right) \pi_k(x) \\
= \left((1 + o(1)) \frac{R_k}{\prod_{q \in Q} (q - 2)} + \frac{1}{\pi_k(x)} \sum_{5 \leq r \leq z} \frac{1}{r - N_f(r)} \log \left(1 + \frac{1}{r - 1}\right)\right) \pi_k(x) \\
< \left(\frac{R_k}{\prod_{q \in Q} (q - 2)} + \frac{\delta}{3}\right) \pi_k(x),
\]

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

(b) Suppose that \( z < r \leq (\log x)^3 \) 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 \( x \),

\[
\pi'_k,r(x) \leq \frac{K(x + \tau_k - b_r)/ar}{(\log((x + \tau_k - b_r)/ar))^2} \prod_{p \geq 2} \frac{1 - N_p(p)/p}{(1 - 1/p)^2} \\
= \frac{C_kK(x + \tau_k - b_r)}{(r - N_f(r))(\log((x + \tau_k - b_r)/ar))^2} \prod_{q \in Q} (q - 2)^{-1} \\
\leq \frac{2C_kKx}{(r - N_f(r))(\log((x - b_r)/ar))^2} \prod_{q \in Q} (q - 2)^{-1} \\
\leq \frac{2C_kKx}{(r - N_f(r))(\log(x/ar - 1))^2} \prod_{q \in Q} (q - 2)^{-1},
\]

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

\[
\log(x/ar - 1) \geq \log(x^{1/2}) = (\log x)/2
\]

for large enough \( x \). Thus,

\[
\pi'_k,r(x) \leq \frac{8C_kKx}{(r - N_f(r))(\log x)^2} \prod_{q \in Q} (q - 2)^{-1} \\
\leq \frac{8K\pi_k(x)}{r - 2} \prod_{q \in Q} (q - 2)^{-1}
\]
for sufficiently large $x$. Since $\log(1 + t) < t$ for $t > 0$, for sufficiently large $x$ we obtain

$$A_2(x) = \sum_{\frac{\log x}{3} < r \leq x, r \notin Q} \pi'_{k, r}(x) \log \left( 1 + \frac{1}{r - 1} \right)$$

$$\leq 8K \pi_k(x) \prod_{q \in Q} (q - 2) \sum_{r > z} \frac{1}{r - 2} \log \left( 1 + \frac{1}{r - 1} \right)$$

$$< 8K \pi_k(x) \prod_{q \in Q} (q - 2) \sum_{r > z} \frac{1}{(r - 2)(r - 1)}$$

$$= 8K \pi_k(x) \prod_{q \in Q} (q - 2) \sum_{r > z} \left( \frac{1}{r - 2} - \frac{1}{r - 1} \right)$$

$$\leq \frac{8K \pi_k(x)}{(z - 2) \prod_{q \in Q} (q - 2)}$$

$$< \frac{\delta}{3} \pi_k(x).$$

(c) Suppose that $(\log x)^3 < r \leq x$ 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

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

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

$$A_3(x) = \sum_{\frac{\log x}{3} < r \leq x, r \notin Q} \pi'_{k, r}(x) \log \left( 1 + \frac{1}{r - 1} \right)$$

$$\leq \sum_{\frac{\log x}{3} < r \leq x, r \notin Q} \frac{1}{r - 1} \left( \frac{x}{ar} + 1 \right)$$

$$\leq \frac{x}{a} \sum_{r > (\log x)^3} \frac{1}{r(r - 1)} + \sum_{\frac{\log x}{3} < r \leq x} \frac{1}{r - 1}$$

$$\leq \frac{x}{a} \sum_{r > (\log x)^3} \left( \frac{1}{r - 1} - \frac{1}{r} \right) + \int_{(\log x)^3 - 2}^{x} \frac{dt}{t}$$

$$\leq \frac{x}{a((\log x)^3 - 1)} + \left( \log t \right)_{t=(\log x)^3 - 2}$$

$$\leq \frac{x}{a((\log x)^3 - 1)} + \log x$$
\[ \pi_k(x) = (1 + o(1)) \left( \frac{x}{\alpha \log^3 x} + \log x \right) \]
\[ = o(1) \pi_k(x) \]
\[ < \frac{\delta}{3} \pi_k(x). \]

Returning to (5.9) and using the preceding three estimates, we have
\[ A(x) = A_1(x) + A_2(x) + A_3(x) \]
\[ < \left( \frac{R_k}{\prod_{q \in Q} (q - 2)} + \delta \right) \pi_k(x) \]
for sufficiently large \( x \). Let \( U(x) \) be the set of primes \( p \) counted by \( \pi'_k(x) \) with \( \text{sgn } S(p) = -\chi_3(k) \), so that \( p \) does not belong to the set of interest in Theorem 7(a). As we have seen, if \( p \in U(x) \), then
\[ F(p) > L_k. \]

Thus,
\[ 0 \leq \#U(x) L_k \]
\[ < \sum_{p \in U(x)} F(p) \leq A(x) \]
\[ < \left( \frac{R_k}{\prod_{q \in Q} (q - 2)} + \delta \right) \pi_k(x), \]
from which we deduce that
\[ \#U(x) < \left( \frac{R_k \prod_{q \in Q} (q - 2)^{-1} + \delta}{L_k} \right) \pi_k(x). \]

The primes \( p \) counted by \( \pi'_k(x) \) which are not in \( U(x) \) satisfy \( \text{sgn } S(p) = \chi_3(k) \). By Lemma 8 and the preceding calculation, for large \( x \) there are at least
\[ \pi'_k(x) - \#U(x) > \left( (1 + o(1)) \prod_{q \in Q} (q - 2)^{-1} - \frac{R_k \prod_{q \in Q} (q - 2)^{-1} + \delta}{L_k} \right) \pi_k(x) \]
\[ = \prod_{q \in Q} (q - 2)^{-1} \left( 1 - \frac{R_k}{L_k} - \epsilon \right) \pi_k(x) \]
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 \( \epsilon \). By Theorem 6, the set of \( p \in \mathbb{P}_k \) for which
\[ \text{sgn } S(p) = \text{sgn } T(p) \]

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

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

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 \( r \geq 5 \) be prime. We wish to count the number of \( p \in \mathbb{P}_k(x) \) for which \( r \mid (p - 1 + \tau_k) \).

If \( r \mid (k - \chi_3(k)) \), 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

\[ g_1(t) = rt + 1 - \tau_k, \quad g_2(t) = rt + 1 + (k - \tau_k), \quad \text{and } g = g_1g_2. \]

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

\[
\sum_{\substack{p \in \mathbb{P}_k(x) \mid p + \tau_k \equiv 1 \pmod{r} \atop r \nmid (k - \chi_3(k))}} 1 = (1 + o(1)) \frac{(x + \tau_k - 1)/r}{(\log((x + \tau_k - 1)/r))^2} \prod_{p \geq 2} \frac{1 - N_g(p)/p}{(1 - 1/p)^2}
\]

\[
= (1 + o(1)) \frac{x}{r(\log x)^2} \cdot \frac{1}{(1 - 1/r)^2} \prod_{p \neq r} \frac{1 - N_g(p)/p}{(1 - 1/p)^2}
\]

\[
= (1 + o(1)) \frac{x}{(\log x)^2} \cdot \frac{1}{r - N_f(r)} \prod_{p \geq 2} \frac{1 - N_f(p)/p}{(1 - 1/p)^2}
\]

\[
= (1 + o(1)) \frac{\pi_k(x)}{r - N_f(r)},
\]

(5.10)
in which $N_f(r)$ refers to (2.2). If $\text{sgn} \ S(p) = \chi_3(k)$, so that $p$ does not belong to the set of interest in Theorem 7(b), then

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

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

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

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

$$\pi_k''(x) \log(3/2) < \sum_{p \in \mathbb{P}_k(x)} G(p)$$

$$= \sum_{p \in \mathbb{P}_k(x)} \sum_{\substack{r | (p - 1 + \tau_k) \\
r \geq 5}} \log \left(1 + \frac{1}{r - 1}\right)$$

$$\leq \sum_{5 \leq r \leq x} \log \left(1 + \frac{1}{r - 1}\right) \sum_{\substack{p \in \mathbb{P}_k(x) \\
p + \tau_k \equiv 1 \mod r}} 1$$

$$\leq \sum_{5 \leq r \leq x} \log \left(1 + \frac{1}{r - 1}\right) \sum_{\substack{p \in \mathbb{P}_k(x) \\
p + k \equiv 1 \mod r}} 1$$

$$+ \sum_{5 \leq r \leq x} \log \left(1 + \frac{1}{r - 1}\right)$$

$$\leq \left[(1 + o(1)) \sum_{r \geq 5} \frac{1}{r - N_f(r)} \log \left(1 + \frac{1}{r - 1}\right) \pi_k'(x) + \frac{1}{\pi_k(x)} \sum_{r \geq 5} \log \left(1 + \frac{1}{r - 1}\right) \right] \pi_k(x)$$

$$= \left(\sum_{r \geq 5} \frac{1}{r - N_f(r)} \log \left(1 + \frac{1}{r - 1}\right) + o(1)\right) \pi_k(x)$$

$$= (R_k' + o(1)) \pi_k(x).$$
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 \( \text{sgn} \, S(p) = -\chi_3(k) \). 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\(^2\)
\[ R'_k = \sum_{r \geq 5} \frac{1}{r - N_f(r)} \log \left( 1 + \frac{1}{r - 1} \right) \]
\[ \leq \sum_{r \geq 5} \frac{1}{r - 2} \log \left( 1 + \frac{1}{r - 1} \right) \]
\[ < 0.1412981. \]

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 \( k \equiv \mp 1 \pmod{3} \), 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 0.1412981112.
\begin{table}[h]
\centering
\caption{Lower bounds from Theorem 10 for $k \equiv 0 \pmod{3}$.}
\begin{tabular}{cccccccc}
\hline
\textbf{$k$} & $\chi_3(k) = -1$ & & $\chi_3(k) = 1$ & \\
 & $Q^+$ & $L_k^-$ & $R_k^-$ & Lwr bd: $T(p) < 0$ & $Q^+$ & $L_k^+$ & $R_k^+$ & Lwr bd: $T(p) > 0$ \\
\hline
6 & 5 & 0.223144 & 0.069917 & 0.233732 & 7 & 0.154151 & 0.110468 & 0.056675 \\
12 & 5 & 0.223144 & 0.056327 & 0.249192 & 5 & 0.223144 & 0.059840 & 0.244212 \\
18 & 5 & 0.223144 & 0.062875 & 0.23941 & 5 & 0.223144 & 0.063773 & 0.238123 \\
24 & 7 & 0.154151 & 0.108351 & 0.050422 & 5 & 0.223144 & 0.066917 & 0.233727 \\
30 & 7 & 0.154151 & 0.090573 & 0.082487 & 7 & 0.154151 & 0.099742 & 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 & 0.064349 & 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 & 0.067111 & 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 & 0.238394 \\
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 & 0.002900 \\
132 & 5 & 0.223144 & 0.065799 & 0.235043 & 5 & 0.223144 & 0.031847 & 0.258756 \\
138 & 5 & 0.223144 & 0.066766 & 0.233597 & 5 & 0.223144 & 0.067683 & 0.235395 \\
144 & 7 & 0.154151 & 0.092601 & 0.079586 & 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 \\
\hline
\end{tabular}
\end{table}

$p + k - 1$, 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

$$R_k^± = \sum_{\substack{r \geq 5 \\
q \in \mathbb{Q}^± \\
r \mid (k ±1)}} \frac{1}{r - N_f(r)} \log \left( 1 + \frac{1}{r - 1} \right),$$

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

$$L_k^± > R_k^±.$$ 

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

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

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

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

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

[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.