J. Aust. Math. Soc. **82** (2007), 133–147

ON VALUES TAKEN BY THE LARGEST PRIME FACTOR OF SHIFTED PRIMES

WILLIAM D. BANKS and IGOR E. SHPARLINSKI✉

(Received 12 April 2005; revised 28 February 2006)

Communicated by W. W. L. Chen

Abstract

Let $\mathcal{P}$ denote the set of prime numbers, and let $P(n)$ denote the largest prime factor of an integer $n > 1$. We show that, for every real number $32/17 < \eta < (4 + 3\sqrt{2})/4$, there exists a constant $c(\eta) > 1$ such that for every integer $a \neq 0$, the set

$$\left\{p \in \mathcal{P} : p = P(q-a) \text{ for some prime } q \text{ with } p^\eta < q < c(\eta)\, p^\eta\right\}$$

has relative asymptotic density one in the set of all prime numbers. Moreover, in the range $2 \leq \eta < (4+3\sqrt{2})/4$, one can take $c(\eta) = 1+\varepsilon$ for any fixed $\varepsilon > 0$. In particular, our results imply that for every real number $0.486 \leq \vartheta \leq 0.531$, the relation $P(q-a) \asymp q^\vartheta$ holds for infinitely many primes q. We use this result to derive a lower bound on the number of distinct prime divisors of the value of the Carmichael function taken on a product of shifted primes. Finally, we study iterates of the map $q \mapsto P(q-a)$ for $a > 0$, and show that for infinitely many primes q, this map can be iterated at least $(\log\log q)^{1+o(1)}$ times before it terminates.

2000 *Mathematics subject classification*: primary 11N25, 11N64.

1. Introduction

1.1. Background Let $\mathcal{P}$ be the set of prime numbers, and for every integer $n > 1$, let $P(n) \in \mathcal{P}$ be the largest prime factor of n. The function $P : \{2, 3, \ldots\} \to \mathcal{P}$ arises naturally in many number theoretic situations and has been the subject of numerous investigations; see, for example, [5, 6, 8, 14, 16, 17, 18, 20, 24, 28] and the references contained therein.

Recently, driven in part by applications to cryptography, there has been a surge of interest in studying the largest prime factors of the 'shifted primes' $\{q \pm 1 : q \in \mathcal{P}\}$.

Improving on earlier results of Pomerance [25], Balog [4], Fouvry and Grupp [12], and Friedlander [13], Baker and Harman [3] proved the existence of infinitely many primes q for which $P(q-a) \le q^{0.2961}$, where $a \neq 0$ is any fixed integer. In the same paper, they also showed the existence of infinitely many primes q for which $P(q-a) \ge q^{0.677}$, improving earlier results of Hooley [19], Deshouillers and Iwaniec [10], Fouvry [11], and others.

In this paper, we study the related problem of estimating the number of primes p that occur as the largest prime factor of a shifted prime $q-a$ when $q \in \mathcal{P}$ lies in a certain interval determined by p. Interestingly, questions of this sort also have applications in theoretical computer science and, in a different form, have been considered by Vishnoi [29].

We also study iterates of the map $q \mapsto P(q-a)$ for $a > 0$, and show that for infinitely many primes q, this map can be iterated at least $(\log\log q)^{1+o(1)}$ times before it terminates.

1.2. Main results For an integer $a \neq 0$ and real numbers $\eta > 0$ and $c > 1$, let $\mathcal{P}_{a,\eta,c}$ be the set of primes:

$$\mathcal{P}_{a,\eta,c} = \left\{p \in \mathcal{P} : p = P(q-a) \text{ for some prime } q \text{ with } p^\eta < q < c\,p^\eta\right\},$$

and let $\pi_{a,\eta,c}(x)$ denote its counting function:

$$\pi_{a,\eta,c}(x) = \#\{p \le x : p \in \mathcal{P}_{a,\eta,c}\}.$$

Denoting by $\pi(x)$, as usual, the number of primes $p \le x$, we show that if η lies in a suitable range, then there exists a constant $c = c(\eta)$ such that

$$\lim_{x\to\infty} \frac{\pi_{a,\eta,c}(x)}{\pi(x)} = 1$$

holds for every integer $a \neq 0$. In other words, $\mathcal{P}_{a,\eta,c}$ has *relative asymptotic density one* in the set of all prime numbers. More precisely, we prove the following.

THEOREM 1.1. *For every real number* $32/17 < \eta < (4+3\sqrt{2})/4$, *there exists a constant* $c = c(\eta) > 1$ *such that the estimate*

$$\pi_{a,\eta,c}(x) = \pi(x) + O\left(\frac{x}{\log^K x}\right),$$

holds for every integer $a \neq 0$ *and real number* K, *where the implied constant depends only on* a, η, *and* K. *Moreover, if* $2 \le \eta < (4+3\sqrt{2})/4$, *this estimate holds for any constant* $c > 1$.

Let $\mathcal{I}_a$ denote the set of all limit points of the set of ratios

$$\left\{ \frac{\log P(q-a)}{\log q} : q \in \mathcal{P} \right\}.$$

Certainly, there is no reason to doubt that $\mathcal{I}_a = [0, 1]$; however, our present knowledge about the structure of $\mathcal{I}_a$ is rather limited. Using the results of [3] mentioned above, it is easy to see that $\inf \mathcal{I}_a \le 0.2961$ and $\sup \mathcal{I}_a \ge 0.677$ for any $a \ne 0$. In view of Theorem 1.1, we immediately deduce the following result.

COROLLARY 1.2. *For every integer $a \ne 0$, the set $\mathcal{I}_a$ contains the closed interval* $[0.486, 0.531]$.

We remark that, under the *Elliott-Halberstam conjecture*, which asserts that the bound

$$\sum_{m \le x^{1-\varepsilon}} \max_{y \le x} \max_{\gcd(a,m)=1} \left| \pi(y; m, a) - \frac{\pi(y)}{\varphi(m)} \right| \ll \frac{x}{\log^C x}$$

holds for any fixed real numbers $\varepsilon, C > 0$, our approach yields an extension of Theorem 1.1 to the range $1 < \eta < (4 + 3\sqrt{2})/4$. The same conjecture also implies that $\mathcal{I}_a = [0, 1]$. Indeed, if $\pi_a(x, y)$ denotes the number of primes $q \le x$ for which $P(q-a) \le y$, then it is natural to expect that the asymptotic relation

$$\pi_a(x, y) \sim \rho(u)\pi(x) \tag{1}$$

holds over a wide range in the xy-plane, where $u = (\log x)/(\log y)$ and $\rho(u)$ is the Dickman function (see [14, 17, 27]). The statement (1) is a well-known consequence of the Elliott-Halberstam conjecture (see [1, 14]), and using (1) it is easy to see that $\mathcal{I}_a = [0, 1]$.

Next, recall that the *Carmichael function* $\lambda(n)$ is defined for $n \ge 1$ as the maximal order of any element in the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^\times$. More explicitly, for a prime power p^ν, one has

$$\lambda(p^\nu) = \begin{cases} p^{\nu-1}(p-1), & \text{if } p \ge 3 \text{ or } \nu \le 2; \\ 2^{\nu-2}, & \text{if } p = 2 \text{ and } \nu \ge 3; \end{cases}$$

and for an arbitrary integer $n \ge 2$,

$$\lambda(n) = \operatorname{lcm}\big(\lambda(p_1^{\nu_1}), \ldots, \lambda(p_k^{\nu_k})\big),$$

where $n = p_1^{\nu_1} \cdots p_k^{\nu_k}$ is the prime factorization of n. Clearly, $\lambda(1) = 1$.

We also use $\omega(n)$, as usual, to denote the number of distinct prime divisors of $n \ge 1$; in particular, $\omega(1) = 0$.

THEOREM 1.3. *For a fixed integer $a \neq 0$, let*

$$Q_a(x) = \prod_{\substack{q \in \mathcal{P} \\ a < q \le x}} (q - a) \quad \text{and} \quad W_a(x) = \omega\big(\lambda(Q_a(x))\big).$$

Then, for sufficiently large x, the lower bound $W_a(x) \ge x^{0.3596}$ holds.

Again, it is an easy matter to verify that, under the Elliott-Halberstam conjecture, the bound $W_a(x) \ge x^{1+o(1)}$ holds for any fixed a.

Now, let $a > 0$ be fixed, and put $\mathcal{Q}_{a,0} = \{q \in \mathcal{P} : q \le a + 1\}$. We define sets of primes $\{\mathcal{Q}_{a,k} : k \ge 1\}$ recursively by

$$\mathcal{Q}_{a,k} = \{q \in \mathcal{P} : q \ge a + 2,\ P(q - a) \in \mathcal{Q}_{a,k-1}\}, \quad k \in \mathbb{N},$$

and consider the corresponding counting functions $\rho_{a,k}(x) = \#\{q \le x : q \in \mathcal{Q}_{a,k}\}$, $k \in \mathbb{N}$.

THEOREM 1.4. *For every integer $a > 0$, the bound*

$$\rho_{a,k}(x) \le 2^a 3^{k+1} x \exp\big(-(\log x)^{1/k}\big)(\log x)^{ak}$$

holds for all $x > x_0(a)$, where $x_0(a)$ depends only on a, and $k \ge 1$.

For fixed $a > 0$ and an arbitrary prime q, consider the chain given by $q_0 = q$, and $q_j = P(q_{j-1} - a)$, $j \in \mathbb{N}$, and define $k_a(q)$ as the smallest nonnegative integer k for which $q_k \le a + 1$.

COROLLARY 1.5. *Let $a > 0$ be fixed. Then for all but $o(\pi(x))$ primes $q \le x$, the following lower bound holds*:

$$k_a(q) \ge (1 + o(1))\frac{\log\log x}{\log\log\log x}.$$

The lower bound of Corollary 1.5 is closely related to (and complements) certain results from [22].

We also observe that $k_1(q)$ gives a lower bound for the height of the tree representing the *Pratt primality certificate* [26] associated to q. This primality certificate is a recursively-defined construction which consists of a primitive root g modulo q and a list of the prime divisors $p_1, \dots, p_s$ of $q - 1$ together with their certificates of primality; accordingly, the whole certificate has the natural structure of a tree. Clearly, the height $H(q)$ of this tree satisfies the trivial bound $H(q) \ll \log q$. On the other hand, our Corollary 1.5 implies that the lower bound

$$H(q) \ge (1 + o(1))\frac{\log\log x}{\log\log\log x} \tag{2}$$

holds for all but $o(\pi(x))$ primes $q \le x$.

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

2. Preliminaries

2.1. Notation Throughout the paper, we adopt the following conventions.

Any implied constants in the symbols O, $\ll$ and $\gg$ may depend (where obvious) on the parameters a, η, and K, but are absolute otherwise. We recall that the statements $A \ll B$ and $B \gg A$ are equivalent to $A = O(B)$ for positive functions A and B.

The letters p, q, r, ℓ are always used to denote prime numbers, and m, n always denote positive integers.

As usual, we write $\pi(x; m, a)$ for the number of primes $p \le x$ in the arithmetic progression $a \pmod m$.

For simplicity, we use $\log x$ to denote the maximum of 1 and the natural logarithm of $x > 0$, and we write $\log_2 x = \log(\log x)$.

Finally, we use $\varphi(n)$ to denote the value of the Euler function at the positive integer n.

2.2. Necessary tools Our principal tool is the following result, which follows immediately from the *Bombieri-Vinogradov theorem* (see [9]) in the range $0 < \vartheta < 1/2$, from [2, Theorem 1] in the range $1/2 \le \vartheta \le 13/25$, and from the main theorem of [23] in the range $13/25 < \vartheta < 17/32$.

LEMMA 2.1. *There exist functions $C_2(\vartheta) > C_1(\vartheta) > 0$, defined for real numbers ϑ in the open interval $(0, 17/32)$, such that for every integer $a \ne 0$ and real number K, the inequalities*

$$\frac{C_1(\vartheta)\, y}{\varphi(p)\log y} < \pi(y; p, a) < \frac{C_2(\vartheta)\, y}{\varphi(p)\log y}$$

hold for all primes $p \le y^{\vartheta}$, with at most $O(y^{\vartheta}/\log^K y)$ exceptions, where the implied constant depends only on a, ϑ, and K. Moreover, for any fixed $\varepsilon > 0$, these functions can be chosen to satisfy the following properties:

- *$C_1(\vartheta)$ is monotonic decreasing, and $C_2(\vartheta)$ is monotonic increasing*;
- *$C_1(1/2) = 1 - \varepsilon$, and $C_2(1/2) = 1 + \varepsilon$.*

We also need a result from sieve theory, which is an application of *Brun's method*. The following statement is Theorem 6.7 of [21] (see also [15, Theorem 5.7]).

LEMMA 2.2. *Let g be a natural number, and let a_j, b_j $(j = 1, \dots, g)$ be integers such that $E \ne 0$, where $E = \prod_{j=1}^{g} a_j \prod_{1\le i<k\le g}(a_i b_k - a_k b_i)$. For a prime number r, let $\rho(r)$ be the number of solutions n modulo r to the congruence*

$$\prod_{j=1}^{g}(a_j n + b_j) \equiv 0 \pmod r,$$

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

and suppose that $\rho(r) < r$ for every r. Then, for $X \ge Y > 1$,

$$\begin{aligned}&\#\{Y - X < n \le X : (a_j n + b_j) \text{ is prime for } j = 1, \dots, g\}\\ &\quad \le 2^g g! \prod_{r\in\mathcal{P}} \left(1 - \frac{\rho(r)}{r}\right)\left(1 - \frac{1}{r}\right)^{-g} \frac{Y}{\log^g Y}\left\{1 + O\left(\frac{\log_2 Y + \log_2 |E|}{\log Y}\right)\right\},\end{aligned}$$

where the implied constant depends only on g.

Finally, we need the following technical result.

LEMMA 2.3. *For every positive integer n, let*

$$\psi(n) = \prod_{\substack{r\in\mathcal{P}\\ r\mid n, r\ne 2}} \left(1 + \frac{1}{r-2}\right).$$

Then the following estimate holds:

$$\sum_{\substack{k\le z\\ \gcd(k,a)=1, 2\mid ka}} \frac{\psi(k)}{k} = \left(c_2^{-1} + o(1)\right) \frac{\varphi(2a)}{2a} \log z \prod_{\substack{r\in\mathcal{P}\\ r\mid a, r\ne 2}} \left(1 - \frac{1}{(r-1)^2}\right),$$

where c_2 is the 'twin primes constant' given by:

$$c_2 = \prod_{\substack{r\in\mathcal{P}\\ r\ne 2}} \left(1 - \frac{1}{(r-1)^2}\right) = 0.6601618158\dots.$$

PROOF. Since $\psi(n) = \sum_{d\mid n, 2\nmid d} \mu^2(d)/F(d)$, where $\mu(d)$ is the Möbius function, and $F(n) = \prod_{r\in\mathcal{P}; r\mid n}(r-2)$, it follows that

$$\sum_{\substack{k\le z\\ \gcd(k,a)=1, 2\mid ka}} \frac{\psi(k)}{k} = \sum_{\substack{k\le z\\ \gcd(k,a)=1, 2\mid ka}} \frac{1}{k} \sum_{\substack{d\mid k\\ 2\nmid d}} \frac{\mu^2(d)}{F(d)} = \sum_{\substack{d\le z\\ \gcd(d,2a)=1}} \frac{\mu^2(d)}{dF(d)} \sum_{\substack{h\le z/d\\ \gcd(h,a)=1, 2\mid ha}} \frac{1}{h}.$$

In the case that a is odd, we have

$$\begin{aligned}\sum_{\substack{h\le z/d\\ \gcd(h,a)=1, 2\mid ha}} \frac{1}{h} &= \sum_{\substack{h\le z/2d\\ \gcd(h,a)=1}} \frac{1}{2h} = \frac{1}{2} \sum_{h\le z/2d} \frac{1}{h} \sum_{\delta\mid \gcd(h,a)} \mu(\delta) = \frac{1}{2} \sum_{\delta\mid a} \frac{\mu(\delta)}{\delta} \sum_{h\le z/2d\delta} \frac{1}{h}\\ &= \frac{1}{2} \sum_{\delta\mid a} \frac{\mu(\delta)}{\delta} (\log(z/d) + O(1)) = \frac{\varphi(a)}{2a} \log(z/d) + O(1),\end{aligned}$$

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

whereas for even a,

$$\sum_{\substack{h\le z/d\\ \gcd(h,a)=1,\,2\mid ha}} \frac{1}{h} = \sum_{\substack{h\le z/d\\ \gcd(h,a)=1}} \frac{1}{h} = \sum_{h\le z/d} \frac{1}{h} \sum_{\delta\mid \gcd(h,a)} \mu(\delta) = \sum_{\delta\mid a} \frac{\mu(\delta)}{\delta} \sum_{h\le z/d\delta} \frac{1}{h}$$
$$= \sum_{\delta\mid a} \frac{\mu(\delta)}{\delta}(\log(z/d) + O(1)) = \frac{\varphi(a)}{a} \log(z/d) + O(1).$$

Hence, in either case, we have

$$\sum_{\substack{k\le z\\ \gcd(k,a)=1,\,2\mid ka}} \frac{\psi(k)}{k} = \frac{\varphi(2a)}{2a} \sum_{\substack{d\le z\\ \gcd(d,2a)=1}} \frac{\mu^2(d)}{dF(d)} \log(z/d) + O(1).$$

Now we split the summation on the right according to whether $d \le w$ or $d > w$, where $w = \exp\left(\sqrt{\log z}\right)$. Since $F(d) \gg \sqrt{d}$ for all odd squarefree integers $d \ge 1$, it follows that

$$\sum_{\substack{w<d\le z\\ \gcd(d,2a)=1}} \frac{\mu^2(d)}{dF(d)} \log(z/d) \ll \frac{\log z}{w^{1/2}} = o(1),$$

and

$$\sum_{\substack{d\le w\\ \gcd(d,2a)=1}} \frac{\mu^2(d)}{dF(d)} \log(z/d) = (1+o(1)) \log z \sum_{\substack{d=1\\ \gcd(d,2a)=1}}^{\infty} \frac{\mu^2(d)}{dF(d)}$$
$$= (1+o(1)) \log z \prod_{\substack{r\in\mathcal{P}\\ r\nmid 2a}} \left(1 + \frac{1}{r(r-2)}\right)$$
$$= (1+o(1))\, c_2^{-1} \log z \prod_{\substack{r\in\mathcal{P}\\ r\mid a,\, r\ne 2}} \left(1 - \frac{1}{(r-1)^2}\right).$$

The result follows. □

3. Proofs

PROOF OF THEOREM 1.1. Let the numbers a, η, and K be fixed as in the statement of Theorem 1.1, and put $\vartheta = 1/\eta$. In what follows, the real numbers $\lambda > 1$ and $\Delta, \mu > 0$ are constants that depend only on a, η, and K.

Let x be a large positive real number, and put $y = x^\eta$. Then

(3) $$x = y^\vartheta \quad \text{and} \quad \log y = \eta \log x.$$

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

Applying Lemma 2.1, first with y and then with λy, we see that the inequalities

$$\frac{C_1(\vartheta)\, y}{\varphi(p)\log y} < \pi(y; p, a) < \frac{C_2(\vartheta)\, y}{\varphi(p)\log y}$$

and

$$\frac{C_1(\vartheta)\, \lambda y}{\varphi(p)\log y} < \pi(\lambda y; p, a) < \frac{C_2(\vartheta)\, \lambda y}{\varphi(p)\log y}$$

hold for all primes $p \le x$, with at most $O\left(x/\log^K x\right)$ exceptions. Hence, if we define the set

$$\mathcal{A} = \left\{ p \le x : \pi(y; p, a) \le \frac{C_2(\vartheta)\, y}{p\log y} \text{ and } \pi(\lambda y; p, a) \ge \frac{C_1(\vartheta)\, \lambda y}{p\log y} \right\},$$

it follows that $\#\mathcal{A} = \pi(x) + O\big(x/\log^K x\big)$.

Next, let $\mathcal{B} = \{p \in \mathcal{A} : p \le (1-\Delta)\,x\}$ and $\mathcal{C} = \{p \in \mathcal{A} : (1-\Delta)\,x < p \le x\}$. Since $\mathcal{A}$ is the disjoint union of $\mathcal{B}$ and $\mathcal{C}$, and

$$\#\mathcal{B} \le \pi((1-\Delta)\,x) = (1-\Delta)\,\pi(x) + O\left(\frac{x}{\log^K x}\right),$$

we see that

$$\#\mathcal{C} \ge \Delta\,\pi(x) + O\left(\frac{x}{\log^K x}\right). \tag{4}$$

For a fixed prime $p \in \mathcal{C}$, let

$$\begin{aligned} \mathcal{D}_p &= \{y < q \le \lambda y : q \equiv a \pmod p \text{ and } P(q-a) > p\},\\ \mathcal{E}_p &= \{y < q \le \lambda y : P(q-a) = p\}, \end{aligned}$$

and observe that

$$\begin{aligned} \#\mathcal{E}_p &= \pi(\lambda y; p, a) - \pi(y; p, a) - \#\mathcal{D}_p \\ &\ge (C_1(\vartheta)\lambda - C_2(\vartheta))\,\frac{y}{p\log y} - \#\mathcal{D}_p. \end{aligned} \tag{5}$$

If $q \in \mathcal{D}_p$, then there exists a prime $\ell > p$ and an integer k such that $q = pk\ell + a$ and $P(q-a) = \ell$. In fact, the condition $P(q-a) = \ell$ is redundant. Indeed, since $\ell > p > (1-\Delta)\,x$, we have $k = (q-a)/p\ell \ll y/x^2 = x^{\eta-2}$, and since $\eta < 3$, it follows that $\ell > p > k$ once x is sufficiently large. Moreover, the preceding estimate implies that $\mathcal{D}_p = \varnothing$ for all $p \in \mathcal{C}$ if $\eta < 2$ and x is large enough.

Next, we estimate $\#\mathcal{D}_p$ in the case that $\eta \ge 2$. For each prime $p \in \mathcal{C}$ and integer $k \ll x^{\eta-2}$, let $\mathcal{F}_{p,k} = \{y < q \le \lambda y : q = pk\ell + a \text{ for some prime } \ell > p\}$. Clearly,

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

$\mathcal{D}_p \subset \bigcup_k \mathcal{F}_{p,k}$. Moreover, assuming that x is large enough, we have $\mathcal{F}_{p,k} \subset \mathcal{G}_{p,k}$, where $\mathcal{G}_{p,k} = \{(1-\mu)y/pk < \ell \le (1+\mu)\lambda y/pk$: both ℓ and $(pk\ell + a)$ are prime$\}$. To estimate $\#\mathcal{G}_{p,k}$, we first observe that $\mathcal{G}_{p,k} = \varnothing$ if either $\gcd(k,a) \ne 1$ or $2 \nmid ka$. On the other hand, if $\gcd(k,a) = 1$ and $2 \mid ka$, then we apply Lemma 2.2 with the choices $g = 2$, $a_1 = 1$, $b_1 = 0$, $a_2 = pk$, and $b_2 = a$. Note that $E = pka \ne 0$. For every prime r, the number $\rho(r)$ of solutions modulo r to the congruence

$$n(pkn + a) \equiv 0 \pmod{r}$$

is *one* if $r \mid pka$, and *two* otherwise; in particular, $\rho(r) < r$ for every prime r. Finally, taking $X = (1+\mu)\lambda y/pk$ and $Y = (\lambda\mu + \lambda + \mu - 1)y/pk$ in the statement of Lemma 2.2 (thus, $X - Y = (1-\mu)y/pk$), we obtain the following bound:

$$\#\mathcal{G}_{p,k} \le 8 \prod_{r \in \mathcal{P}} \left(1 - \frac{\rho(r)}{r}\right)\left(1 - \frac{1}{r}\right)^{-2} \frac{\gamma y}{pk \log^2(\gamma y/pk)} \times \left\{1 + O\left(\frac{\log_2(\gamma y/pk) + \log_2(pka)}{\log(\gamma y/pk)}\right)\right\},$$

where $\gamma = \lambda\mu + \lambda + \mu - 1$. Noting that $\gamma > \lambda - 1 > 0$, and using the simple estimates $pka \ll x^{\eta - 1}$ and $\gamma y/pk \gg x$, we deduce that

$$\#\mathcal{G}_{p,k} \le (8\gamma + o(1)) \prod_{r \in \mathcal{P}} \left(1 - \frac{\rho(r)}{r}\right)\left(1 - \frac{1}{r}\right)^{-2} \frac{y}{pk \log^2 x}.$$

Now, since p, k, and a are pairwise coprime, and $2 \mid ka$, it follows that

$$\prod_{r \in \mathcal{P}} \left(1 - \frac{\rho(r)}{r}\right)\left(1 - \frac{1}{r}\right)^{-1} = 2c_2 \psi(p)\psi(k)\psi(a),$$

where the constant c_2 and the function $\psi(k)$ are defined as in Lemma 2.3. Therefore,

$$\#\mathcal{G}_{p,k} \le (16 c_2 \gamma + o(1))\, \psi(p)\psi(k)\psi(a) \frac{y}{pk \log^2 x}.$$

Summing this estimate over k, applying Lemma 2.3, and using the fact that $\psi(p) = (1 + o(1))$, we derive that

$$\begin{aligned}\#\mathcal{D}_p &\le \sum_{\substack{k \ll x^{\eta-2} \\ \gcd(k,a)=1,\, 2 \mid ka}} \#\mathcal{G}_{p,k} \le (16 c_2 \gamma + o(1))\, \psi(a) \frac{y}{p \log^2 x} \sum_{\substack{k \ll x^{\eta-2} \\ \gcd(k,a)=1,\, 2 \mid ka}} \frac{\psi(k)}{k} \\ &= (8\gamma(\eta - 2) + o(1)) \frac{y}{p \log x} \frac{\psi(a)\varphi(2a)}{a} \prod_{\substack{r \in \mathcal{P} \\ r \mid a,\, r \ne 2}} \left(1 - \frac{1}{(r-1)^2}\right).\end{aligned}$$

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

It is easy to verify that, for every integer $a \neq 0$, one has

$$\frac{\psi(a)\varphi(2a)}{a} \prod_{\substack{r \in \mathcal{P} \\ r \mid a, r \neq 2}} \left(1 - \frac{1}{(r-1)^2}\right) = 1,$$

and therefore, using (3), we have

$$\#\mathcal{D}_p \leq (8\gamma(\eta - 2) + o(1)) \frac{y}{p \log x} = (8\gamma\eta(\eta - 2) + o(1)) \frac{y}{p \log y}. \tag{6}$$

We now turn to the selection of the constants $c, \lambda > 1$ and $\Delta, \mu > 0$. Our first goal is to show, for x sufficiently large, that $\mathcal{C} \subset \mathcal{P}_{a,\eta,c}$. Suppose that $p \in \mathcal{C}$ and $q \in \mathcal{E}_p$. Then,

$$p^\eta \leq x^\eta = y < q \leq \lambda y = \lambda x^\eta < \frac{\lambda}{(1-\Delta)^\eta} p^\eta.$$

From these inequalities, it follows that $\mathcal{C} \subset \mathcal{P}_{a,\eta,c}$ if $\mathcal{E}_p \neq \varnothing$ for all $p \in \mathcal{C}$, and the constants c, λ, and Δ satisfy the relation

$$c = \frac{\lambda}{(1-\Delta)^\eta}. \tag{7}$$

In the case that $\eta < 2$, we have already seen that $\mathcal{D}_p = \varnothing$ for all $p \in \mathcal{C}$ once x is large enough. By (5), it follows that $\mathcal{E}_p \neq \varnothing$ for all $p \in \mathcal{C}$ if

$$\lambda > \frac{C_2(\vartheta)}{C_1(\vartheta)}. \tag{8}$$

Since $\vartheta > 0.5$ in this case, Lemma 2.1 implies that for any $c \geq C_2(\vartheta)/C_2(\vartheta)$, both relations (7) and (8) can be simultaneously satisfied for an appropriate choice of $\lambda > 1$ and $\Delta > 0$, provided that $\eta > 32/17$.

In the case that $\eta \geq 2$, after substituting (6) into (5), we see that $\mathcal{E}_p \neq \varnothing$ for all $p \in \mathcal{C}$ if x is sufficiently large, and

$$C_1(\vartheta)\lambda - C_2(\vartheta) > 8\gamma\eta(\eta - 2). \tag{9}$$

Let $\varepsilon > 0$ be a fixed constant, and put $c = 1 + \varepsilon$. Choosing

$$\lambda = 1 + \varepsilon/2 > 1 \quad \text{and} \quad \Delta = 1 - \left(\frac{1 + \varepsilon/2}{1 + \varepsilon}\right)^\vartheta > 0,$$

we see that relation (7) is satisfied. For any fixed constant $\delta > 0$, we can assume $C_1(\vartheta) = 1 - \delta$ and $C_2(\vartheta) = 1 + \delta$ according to Lemma 2.1. Since the left-hand side of (9) and $\gamma = \lambda\mu + \lambda + \mu - 1$ can each be made arbitrarily close to $\lambda - 1 = \varepsilon/2$ by choosing δ and μ sufficiently close to 0, it follows that relation (8) is also satisfied for an appropriate choice of δ and μ, provided that $8\eta(\eta-2) < 1$, that is, $\eta < (4+3\sqrt{2})/4$.

Taking into account (4), we have therefore shown that if $32/17 < \eta < (4+3\sqrt{2})/4$, there exists a constant $c = c(\eta)$ such that at least $\Delta\,\pi(x) + O\big(x/\log^K x\big)$ primes p in the interval $(1-\Delta)x < p \le x$ lie in the set $\mathcal{P}_{a,\eta,c}$, and the theorem follows. $\square$

PROOF OF THEOREM 1.3. For fixed η in the range $32/17 < \eta < (4+3\sqrt{2})/4$, put $\vartheta = 1/\eta$, and consider the counting function

$$\varpi_a(y) = \#\{p \le y : p = P(q-a) \text{ for some } q \in \mathcal{P} \text{ with } q \le c\,p^\eta\},$$

where $c = c(\eta)$ is the constant described in Theorem 1.1. According to that theorem, we have the following estimate:

$$\varpi_a(y) = \pi(y) + O\left(\frac{y}{\log^2 y}\right).$$

Defining $y = (c^{-1}x)^\vartheta$, it follows that there are $(1+o(1))\pi(y)$ primes $p \le y$ such that $p \mid Q_a(x)$; let $\mathcal{S}$ denote this set of primes. By a result of [3], there are at least $y^{1+o(1)}$ primes $p \in \mathcal{S}$ with $P(p-a) \ge y^{0.677}$; let $\mathcal{R}$ denote this set of primes. Then

$$W_a(x) \ge \omega\left(\lambda\left(\prod_{p\in\mathcal{S}} p\right)\right) \ge \omega\left(\lambda\left(\prod_{p\in\mathcal{R}} p\right)\right).$$

Let $\mathcal{L}$ be the set of primes ℓ for which $\ell = P(p-1)$ for some $p \in \mathcal{R}$.

Clearly, a prime $\ell \ge y^{0.677}$ cannot have the property that $\ell = P(p-1)$ for more than $y^{0.323}$ primes $p \in \mathcal{R}$. Consequently, $W_a(x) \ge \#\mathcal{L} \ge y^{0.677+o(1)} = x^{0.677\vartheta+o(1)}$. Taking $\eta = \vartheta^{-1}$ sufficiently close to $32/17$, we obtain the stated result. $\square$

PROOF OF THEOREM 1.4. Let $\psi(x,y) = \#\{n \le x : P(n) \le y\}$. We recall the well known bound (see [7, 17, 27])

$$\psi(x,y) \le x\exp(-(1+o(1))u\log u), \tag{10}$$

which holds uniformly as $u = (\log x)/(\log y) \to \infty$ with $u \le y^{1/2}$.

If $k = 1$, then the set $\mathcal{Q}_{a,1} \cap [1,x]$ consists of all prime numbers q of the form $m+a$, where $P(m) \le a+1$ and $m \le x-a < x$. Therefore, the bound

$$\rho_{a,1}(x) \le (2\log x)^{\pi(a+1)} \le (2\log x)^a$$

holds uniformly for $x \ge 2$.

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

We finish the proof by induction. Suppose that the result is true up to $k-1$, where $k \geq 2$. We can assume that

$$k \leq \frac{\log\log x}{\log\log\log x}, \tag{11}$$

for otherwise the bound of the theorem holds trivially. For every integer $w \geq 0$, we have:

$$\begin{aligned}\rho_{a,k}(x) &\leq \psi(x, e^w) + \sum_{\substack{q \in \mathcal{Q}_{a,k-1} \\ e^w < q \leq x}} \lfloor x/q \rfloor \leq \psi(x, e^w) + x \sum_{\substack{q \in \mathcal{Q}_{a,k-1} \\ e^w < q \leq x}} \frac{1}{q} \\ &\leq \psi(x, e^w) + x \sum_{\nu=w}^{\lfloor \log x \rfloor} e^{-\nu} \sum_{\substack{q \in \mathcal{Q}_{a,k-1} \\ e^\nu < q \leq e^{\nu+1}}} 1 \leq \psi(x, e^w) + x \sum_{\nu=w}^{\lfloor \log x \rfloor} e^{-\nu} \rho_{a,k-1}(e^{\nu+1}).\end{aligned}$$

We now choose $u = (\log x)^{1/k}$, and put $w = (\log x)/u = (\log x)^{1-1/k}$. Since $u \geq \log\log x$ by (11), and $u \leq e^{w/2}$ if x is large enough (independent of k or a), we can use the bound (10). In a weaker form, this gives

$$\psi(x, e^w) \leq x e^{-u} = x \exp\left(-(\log x)^{1/k}\right)$$

if x is sufficiently large.

Using the inductive hypothesis, we derive that

$$\begin{aligned}\sum_{\nu=w}^{\lfloor \log x \rfloor} e^{-\nu} \rho_{a,k-1}(e^{\nu+1}) &\leq \sum_{\nu=w}^{\lfloor \log x \rfloor} 2^a 3^k e \exp\left(-(\nu+1)^{1/(k-1)}\right)(\log x)^{a(k-1)} \\ &\leq 2^a 3^k e \exp\left(-w^{1/(k-1)}\right)(\log x)^{a(k-1)+1}.\end{aligned}$$

Therefore

$$\rho_{a,k}(x) \leq x \exp\left(-(\log x)^{1/k}\right) + 2^a 3^k e x \exp\left(-w^{1/(k-1)}\right)(\log x)^{a(k-1)+1}.$$

Since $w^{1/(k-1)} = (\log x)^{1/k}$, $a \geq 1$, and $1 + 3^k e \leq 3^{k+1}$, we conclude the proof. □

4. Concluding remarks

As we have already remarked, the Elliott-Halberstam conjecture leads to an extension of Theorem 1.1 to the range $1 < \eta < (4+3\sqrt{2})/4$. We also note that the factor $2^g g!$ in Lemma 2.2 is probably unnecessary. In the absence of this factor, the stronger estimate of Lemma 2.2 would lead to a corresponding extension of Theorem 1.1 to the range $32/17 < \eta < 1+\sqrt{2}$.

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

Clearly, Theorem 1.3 implies that $\lambda\left(\lambda\left(\prod_{q\in\mathcal{P},\ q\le x} q\right)\right) \gg \exp(x^{0.3596})$. It also would be interesting to estimate k-fold iterates of the Carmichael function applied to the product of the primes $q \le x$.

We now recall the asymptotic formula

$$\frac{1}{x}\sum_{2\le n\le x}\frac{\log P(n)}{\log n} = 0.6243\ldots + o(1)$$

for the average logarithmic size of the largest prime factor (see [27, Exercise 3, Chapter III.5]). Assuming that the shifts $q_j - a$, where $q_0 = q$, and $q_j = P(q_{j-1} - a)$, $j = 1, 2, \ldots$, behave as 'typical' integers, then it is reasonable to expect that the bound $k_a(q) \ll \log\log q$ holds for almost all primes q. In particular, the lower bound of Corollary 1.5 is probably rather tight. On the other hand, it should be possible to improve the logarithmic factor $(\log x)^k$ in the bound of Theorem 1.1 and thus obtain a slightly better bound for $k_a(q)$, although the technical details are more involved. Similarly, although we expect that Theorem 1.1 and Corollary 1.5 also hold for negative integers a (with appropriate modifications), the proof appears to be more complicated as the induction step must be handled in a different way to retain uniformity of the bound with respect to k.

Finally, it would be interesting to know whether the lower bound (2) for the height of the Pratt tree is tight.

Acknowledgements

The authors would like to thank Carl Pomerance for many useful comments and, in particular, for the observation that the Elliott-Halberstam conjecture implies $\mathcal{I}_a = [0, 1]$.

We also thank Kevin Ford for bringing to our attention the link between our results and the Pratt primality certificate.

This work was done during a visit by the second author to the University of Missouri-Columbia; the support and hospitality of this institution are gratefully acknowledged. During the preparation of this paper, the second author was supported in part by ARC Grant DP0556431.

References

[1] W. R. Alford, A. Granville and C. Pomerance, 'There are infinitely many Carmichael numbers', *Annals Math.* **140** (1994), 703–722.

https://doi.org/10.1017/S1446788700017511 Published online by Cambridge University Press

[2] R. C. Baker and G. Harman, 'The brun-titchmarsh theorem on average', in: *Proc. Conf. in Honor of Heini Halberstam (Allerton Park, IL, 1995)*, Progr. Math. 138 (Birkhäuser, Boston, 1996) pp. 39–103.

[3] ———, 'Shifted primes without large prime factors', *Acta Arith.* **83** (1998), 331–361.

[4] A. Balog, '$p + a$ without large prime factors', in: *Seminar on Number Theory, 1983–84 Talence* (Univ. Bordeaux, Thalence, 1984).

[5] W. Banks, J. B. Friedlander, C. Pomerance and I. E. Shparlinski, 'Multiplicative structure of values of the Euler function', in: *High Primes and Misdemeanours: Lectures in Honour of the 60th Birthday of Hugh Cowie Williams*, Fields Institute Communications 41 (Amer. Math. Soc., Providence, RI, 2004) pp. 29–48.

[6] W. Banks, G. Harman and I. E. Shparlinski, 'Distributional properties of the largest prime factor', *Michigan Math. J.* **53** (2005), 665–681.

[7] E. R. Canfield, P. Erdős and C. Pomerance, 'On a problem of Oppenheim concerning "Factorisatio Numerorum"', *J. Number Theory* **17** (1983), 1–28.

[8] C. Dartyge, G. Martin and G. Tenenbaum, 'Polynomial values free of large prime factors', *Periodica Math. Hungar.* **43** (2001), 111–119.

[9] H. Davenport, *Multiplicative number theory*, 2nd edition (Springer, New York, 1980).

[10] J.-M. Deshouillers and H. Iwaniec, 'On the Brun-Titchmarsh theorem on average', in: *Topics in classical number theory, Vol. I, II (Budapest, 1981)*, Colloq. Math. Soc. János Bolyai 34 (North-Holland, Amsterdam, 1984) pp. 319–333.

[11] E. Fouvry, 'Théorème de Brun-Titchmarsh; application au théorème de Fermat', *Invent. Math.* **79** (1985), 383–407.

[12] E. Fouvry and F. Grupp, 'On the switching principle in sieve theory', *J. Reine Angew. Math.* **370** (1986), 101–126.

[13] J. B. Friedlander, *Shifted primes without large prime factors*, Number Theory and Applications 1989 (Kluwer, Berlin, 1990) pp. 393–401.

[14] A. Granville, 'Smooth numbers: Computational number theory and beyond', in: *Algorithmic Number Theory* (eds. J. Buhler and P. Stevenhagen), MSRI Publications 44 (Cambridge Univ. Press, to appear).

[15] H. Halberstam and H.-E. Richert, *Sieve Methods* (Academic Press, London, 1974).

[16] A. Hildebrand, 'On the number of positive integers $\leq x$ and free of prime factors $> y$', *J. Number Theory* **22** (1986), 289–307.

[17] A. Hildebrand and G. Tenenbaum, 'Integers without large prime factors', *J. Théor. Nombres Bordeaux* **5** (1993), 411–484.

[18] N. A. Hmyrova, 'On polynomials with small prime divisors, II', *Izv. Akad. Nauk SSSR Ser. Mat.* **30** (1966), 1367–1372 (in Russian).

[19] C. Hooley, 'On the largest prime factor of $p + 1$', *Mathematika* **20** (1973), 135–143.

[20] A. Ivić, 'On sums involving reciprocals of the largest prime factor of an integer, II', *Acta Arith.* **71** (1995), 229–251.

[21] H. Iwaniec and E. Kowalski, *Analytic number theory* (Amer. Math. Soc., Providence, RI, 2004).

[22] F. Luca and C. Pomerance, 'Irreducible radical extensions and the Euler-function chains', Preprint, 2005.

[23] H. Mikawa, 'On primes in arithmetic progressions', *Tsukuba J. Math.* **25** (2001), 121–153.

[24] S.-M. Oon, 'Pseudorandom properties of prime factors', *Period. Math. Hungar.* **49** (2004), 45–63.

[25] C. Pomerance, 'Popular values of euler's function', *Mathematika* **27** (1980), 84–89.

[26] V. Pratt, 'Every prime has a succinct certificate', *SIAM J. Comput.* **4** (1975), 214–220.

[27] G. Tenenbaum, *Introduction to analytic and probabilistic number theory* (Cambridge University Press, 1995).

[28] N. M. Timofeev, *Polynomials with small prime divisors*, Taškent. Gos. Univ., Naučn. Trudy 548,

Voprosy Mat. (Taškent. Gos. Univ., Taškent, 1977) pp. 87–91 (Russian).
[29] N. K. Vishnoi, *Theoretical aspects of randomization in computation* (Ph.D. Thesis, Georgia Inst. of Technology, 2004), available from
`http://smartech.gatech.edu:8282/dspace/handle/1853/5049`.

Department of Mathematics
University of Missouri
Columbia, MO 65211
USA
e-mail: bbanks@math.missouri.edu

Department of Computing
Macquarie University
Sydney, NSW 2109
Australia
e-mail: igor@ics.mq.edu.au