COLLOQUIUM MATHEMATICUM

VOL. 143 2016 NO. 1

SHIFTED VALUES OF THE LARGEST PRIME FACTOR FUNCTION AND ITS AVERAGE VALUE IN SHORT INTERVALS

BY

JEAN-MARIE DE KONINCK (Québec) and IMRE KÁTAI (Budapest)

Abstract. We obtain estimates for the average value of the largest prime factor $P(n)$ in short intervals $[x, x+y]$ and of $h(P(n)+1)$, where h is a complex-valued additive function or multiplicative function satisfying certain conditions. Letting $s_q(n)$ stand for the sum of the digits of n in base $q \geq 2$, we show that if α is an irrational number, then the sequence $(\alpha s_q(P(n)))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

1. Introduction and notation. Let $P(n)$ stand for the largest prime factor of an integer $n \geq 2$ and set $P(1) = 1$. This function has been extensively studied over the past decades, in particular its average value, sums involving the reciprocals of its values, as well as its most frequent value in the interval $[2, x]$.

Here, we obtain estimates for $\sum_{x\leq n\leq x+y} P(n)$ when $y = x^{7/12+\varepsilon}$ for any $0 < \varepsilon < 5/12$. Given an integer $a \neq 0$, we also obtain estimates for the average value of $h(P(n) + a)$ for various arithmetic functions h satisfying certain regularity conditions. Letting $s_q(n)$ stand for the sum of the digits of n in base $q \geq 2$, we show that if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, the sequence $(\alpha s_q(P(n)))_{n\in\mathbb{N}}$ is uniformly distributed modulo 1.

Before we state these results more explicitly, we provide some background results.

In 1984, De Koninck and Ivić [4] proved that, for every positive integer m, there exist computable constants $d_1 = \pi^2/12, d_2, \ldots, d_m$ such that

$$\sum_{n\leq x} P(n) = x^2\left(\frac{d_1}{\log x} + \frac{d_2}{\log^2 x} + \cdots + \frac{d_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x}\right)\right). \tag{1.1}$$

Recently, Naslund [21] improved (1.1) by showing that, given any $\varepsilon > 0$,

---

2010 *Mathematics Subject Classification*: 11K06, 11N37.

Key words and phrases: largest prime factor, shifted prime.
Received 13 December 2014; revised 22 July 2015.
Published online 3 December 2015.

DOI: 10.4064/cm6474-12-2015 © Instytut Matematyczny PAN, 2016

there exists a positive constant c such that

$$\sum_{n\le x} P(n) = x\,\mathrm{li}_g(x) + O_\varepsilon(x^2\exp\{-c(\log x)^{3/5-\varepsilon}\}),$$

where

$$\mathrm{li}_g(x) = \int_2^x \frac{t}{x}\,\frac{\lfloor x/t\rfloor}{\log t}\,dt = \frac{c_0 x}{\log x} + \frac{c_1 x}{\log^2 x} + \cdots + \frac{c_{m-1}(m-1)!x}{\log^m x} + O\left(\frac{x}{\log^{m+1} x}\right)$$

(for any given $m\in\mathbb{N}$) with the constants c_i defined by

$$c_i = \frac{1}{2^{i+1}}\sum_{j=0}^{i}\frac{2^j(-1)^j\zeta^{(j)}(2)}{j!},$$

where ζ stands for the Riemann Zeta Function.

In 1986, Erdős, Ivić and Pomerance [11] proved that

$$\sum_{n\le x}\frac{1}{P(n)} = x\delta(x)\left(1 + O\left(\sqrt{\frac{\log\log x}{\log x}}\right)\right),$$

where $\delta(x)$ is some continuous function which decreases to 0 very slowly as $x\to\infty$ and in fact satisfies

$$\delta(x) = \exp\{-(1+o(1))\sqrt{2\log x\log\log x}\} \quad \text{as } x\to\infty.$$

On the other hand, it is known (see De Koninck and Luca [5, Problem 9.33]) that

$$\sum_{2\le n\le x}\log P(n) = \kappa x\log x + O(x\log\log x), \tag{1.2}$$

where

$$\kappa = 1 - \int_1^\infty \frac{\rho(v)}{v^2}\,dv \tag{1.3}$$

with $\rho(v)$ standing for the Dickman function.

In 1987, De Koninck and Sitaramachandrarao [6] proved that

$$\sum_{2\le n\le x}\frac{1}{n\log P(n)} = e^\gamma\log\log x + O(1),$$

where γ stands for the Euler–Mascheroni constant.

In 1994, the first author [3], and later De Koninck and Sweeney [7], studied the function

$$f(x,p) := \#\{n\le x : P(n) = p\} \tag{1.4}$$

and proved in particular that the maximum value of $f(x,p)$, as p runs over the interval $[2,x]$, is reached at

$$p = \exp\left\{\sqrt{\frac{1}{2}\log x \log\log x}\left(1 + \lambda(x) + o\left(\frac{1}{\log\log x}\right)\right)\right\} \qquad (x \to \infty),$$

where $\lambda(x) = \frac{1}{2}\frac{\log\log\log x}{\log\log x}$, in which case $f(x,p)$ is equal to

$$x \exp\left\{-\sqrt{2\log x \log\log x}\left(1 + \frac{\lambda(x)}{2} - \frac{2 + \log 2 + o(1)}{2\log\log x}\right)\right\} \qquad (x \to \infty).$$

Some improvements of this result have been obtained by McNew [20].

From now on, we shall write $\pi(x)$ for the number of primes $p \le x$, and $\pi(x;k,\ell)$ for the number of primes $p \equiv \ell \pmod k$ not exceeding x. Moreover, we let $\wp$ stand for the set of all primes.

Now, given a real-valued additive function g such that the set $\{g(p) : p \in \wp\}$ is bounded, let

$$A_x := \sum_{p \le x} \frac{g(p)}{p} \quad \text{and} \quad B_x^2 := \sum_{p \le x} \frac{g^2(p)}{p},$$

and further set

$$\kappa_n := \frac{g(n) - A_n}{B_n} \quad (n \in \mathbb{N}) \quad \text{and} \quad \Phi(u) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-w^2/2}\, dw \quad (u \in \mathbb{R}).$$

According to the Erdős–Kac Theorem (see Elliott [9, Theorem 12.3]), if $B_x \to \infty$ as $x \to \infty$, then

$$\lim_{x\to\infty} \frac{1}{x}\#\{n \le x : \kappa_n < u\} = \Phi(u) \qquad \text{for every real } u.$$

Given a positive integer N, let $\wp_N := \{p \le N : p \in \wp\}$. We shall say that $\rho_N : \wp_N \to [0,1)$ is a *prime weight function* if it satisfies the following four conditions:

(i) $\sum_{p \in \wp_N} \rho_N(p) = 1 + o(1)$ as $N \to \infty$;

(ii) for every non-increasing sequence $(\lambda_N)_{N\in\mathbb{N}}$ tending to 0 as $N \to \infty$,

$$\sum_{\substack{p < N^{\lambda_N}\\ p \in \wp_N}} \rho_N(p) \to 0 \quad \text{and} \quad \sum_{\substack{N^{1-\lambda_N} < p < N\\ p \in \wp_N}} \rho_N(p) \to 0 \qquad (N \to \infty);$$

(iii) with $(\lambda_N)_{N\in\mathbb{N}}$ as in (ii),

$$\max_{\substack{N^{\lambda_N} < p_1 < p_2 < 2p_1 < N^{1-\lambda_N}\\ p_1, p_2 \in \wp_N}} \left|\frac{\rho_N(p_1)}{\rho_N(p_2)} - 1\right| \to 0 \qquad \text{as } N \to \infty;$$

(iv) $\sup_{H \le N} \left|\sum_{H \le p < 2H,\, p \in \wp_N} \rho_N(p)\right| \to 0$ as $N \to \infty$.

An example of a weight function is

$$\rho_N(p) := \frac{c_0}{p}\exp\Big\{-\frac{\log N}{\log p}\Big\}, \qquad \text{where } c_0 = \Big(\int_1^\infty e^{-v}\,\frac{dv}{v}\Big)^{-1}.$$

Indeed, in this case,

$$\begin{aligned}\sum_{p\le N}\rho_N(p) &= c_0\sum_{p\le N}\frac{1}{p}\exp\Big\{-\frac{\log N}{\log p}\Big\} = c_0(1+o(1))\int_2^N \frac{1}{t\log t}e^{-\log N/\log t}\,dt\\ &= c_0(1+o(1))\int_1^{\log N/\log 2} e^{-v}\,\frac{dv}{v}\\ &= c_0(1+o(1))\Big(\int_1^\infty e^{-v}\,\frac{dv}{v} + O\Big(\frac{1}{N\log N}\Big)\Big)\\ &= c_0(1+o(1))\Big(\frac{1}{c_0} + O\Big(\frac{1}{N\log N}\Big)\Big) = 1+o(1),\end{aligned}$$

so that (i) is satisfied. Conditions (ii)–(iv) are also easily verified.

It is known (see [9, Theorem 12.4]) that, under the conditions of the Erdős–Kac Theorem, for every $a \in \mathbb{Z}\setminus\{0\}$,

$$\lim_{N\to\infty} \#\{p \in \wp_N : \kappa_{p+a} < u\} = \Phi(u) \qquad \text{for every real } u$$

and

$$\lim_{N\to\infty} \sum_{\kappa_{p+a}<u} \rho_N(p) = \Phi(u) \qquad \text{for every real } u.$$

According to the Erdős–Wintner Theorem (see Elliott [8, Theorem 5.1]), in order for a real-valued additive function g to have a limiting distribution, it is both sufficient and necessary that it satisfies the *three-series condition*

$$\sum_{|g(p)|\ge 1}\frac{1}{p} < \infty, \qquad \sum_{|g(p)|<1}\frac{g(p)}{p} \text{ converges}, \qquad \sum_{|g(p)|<1}\frac{g^2(p)}{p} < \infty. \tag{1.5}$$

In 1968, the second author [16] proved that if g is a real-valued additive function and

$$F_x(y) := \frac{1}{\mathrm{li}(x)}\sum_{\substack{p\le x\\ g(p+1)<y}} 1, \quad \text{where} \quad \mathrm{li}(x) := \int_2^x \frac{dt}{\log t},$$

and if moreover g satisfies (1.5), then the distribution function $F_x(y)$ tends to a limiting distribution function $F(y)$ as $x\to\infty$ at all points of continuity of $F(y)$. The same holds for $g(p+a)$ for any $a\in\mathbb{Z}\setminus\{0\}$).

Erdős and Kubilius asked whether the three-series condition is necessary or not in the case of shifted primes. In fact, partial results were achieved by

Elliott [10], Kátai [17] and Timofeev [24]. In the end, Hildebrand [13] proved the necessity of the three-series condition for shifted primes.

Now, letting

$$Q_{\mathrm{pr}}(x) = \frac{1}{\pi(x)} \sup_{h\in\mathbb{R}} \#\{p \le x : g(p+a) \in [h, h+1]\}, \tag{1.6}$$

going back to an idea of Ruzsa [23], Timofeev [24] proved that

$$Q_{\mathrm{pr}}(x) \le c\frac{\log^2(2+W(x))}{\sqrt{W(x)}}, \tag{1.7}$$

where

$$W(x) := \min_{\lambda}\biggl(\lambda^2 + \sum_{p\le x} \frac{1}{p} \min\bigl(1, (g(p) - \lambda \log p)^2\bigr)\biggr). \tag{1.8}$$

Later, Elliott [10] refined (1.7) to

$$Q_{\mathrm{pr}}(x) \ll W(x)^{-1/2}.$$

Let $\tau(n)$ be the number of positive divisors of n. Using his *dispersion method*, Linnik [18] proved in 1963 that there exists a constant $d_0 > 0$ such that

$$\sum_{p\le x} \tau(p-1) = d_0 x + O\biggl(\frac{x}{\log^c x}\biggr), \tag{1.9}$$

where $c = 0.999$. Later, in 1986, Bombieri, Friedlander and Iwaniec [1], and independently Fouvry [12], improved (1.9) by showing that, given any $A > 0$ and any integer $a \ne 0$,

$$\sum_{p\le x} \tau(p+a) = D_a x + 2E_a \operatorname{li}(x) + O\biggl(\frac{x}{\log^A x}\biggr), \tag{1.10}$$

where

$$D_a = \frac{\zeta(2)\zeta(3)}{\zeta(6)} \prod_{p\mid a}\biggl(1 - \frac{p}{p^2-p+1}\biggr), \tag{1.11}$$

$$E_a = D_a\biggl(\gamma - \sum_{p} \frac{\log p}{p^2-p+1} + \sum_{p\mid a} \frac{p^2 \log p}{(p-1)(p^2-p+1)}\biggr).$$

On the other hand, if $r(n)$ is the number of representations of the positive integer n as a sum of two squares, it was proved by Hooley [15] that, given any $a \in \mathbb{Z}\setminus\{0\}$ and assuming the General Riemann Hypothesis (GRH),

$$\sum_{p\le x} r(p+a) = (R_a + o(1)) \operatorname{li}(x) \qquad (x \to \infty) \tag{1.12}$$

for a certain positive constant R_a. Later Bredikhin [2] proved (1.12) without assuming GRH; he used the Linnik dispersion method.

Given an integer $q \ge 2$, let $s_q(n)$ be the sum of the digits of n in base q. Mauduit and Rivat [19] proved that:

(i) there exists a constant $\sigma_q(\alpha) > 0$ such that

$$\sum_{n \le x} \Lambda(n) e(\alpha s_q(n)) = O_{q,\alpha}(x^{1-\sigma_q(\alpha)}),$$

where Λ stands for the von Mangoldt function;

(ii) given an integer $m \ge 2$ and setting $d = (q-1, m)$, there exists a constant $\sigma_{q,m} > 0$ such that for every $a \in \mathbb{Z} \setminus \{0\}$, we have

$$\#\{p \le x : s_q(p) \equiv a \pmod{m}\} = \frac{d}{m}\pi(x; d, a) + O_{q,m}(x^{1-\sigma_{q,m}});$$

(iii) the sequence $(\alpha s_q(p))_{p \in \wp}$ is uniformly distributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

In what follows, the letters c and C stand for positive constants, not necessarily the same at each occurrence.

2. Main results

Theorem 1. *Let $f : \wp \to \mathbb{C}$ be a bounded function. Assume that for some constant $\eta \in \mathbb{C}$,*

$$S(x) := \sum_{p \le x} f(p) = (\eta + o(1))\pi(x) \quad (x \to \infty). \tag{2.1}$$

Then

$$\sum_{p \le N} f(p)\rho_N(p) \to \eta \quad (N \to \infty).$$

Theorem 2. *Let g be a real-valued additive function. Then the function $g(P(n)+1)$ has a limiting distribution if and only if g satisfies the three-series condition* (1.5).

Theorem 3. *Let $a \in \mathbb{Z} \setminus \{0\}$. Then*

$$\sum_{n \le x} \tau(P(n) + a) = (\kappa D_a + o(1)) x \log x \quad (x \to \infty),$$

where κ and D_a are the constants defined in (1.3) *and* (1.11)*, respectively.*

Theorem 4. *Let $a \in \mathbb{Z} \setminus \{0\}$. Then*

$$\sum_{n \le x} r(P(n) + a) = (\kappa R_a + o(1)) x \quad (x \to \infty),$$

where R_a is the constant appearing in (1.12).

THEOREM 5. *Let* $y = x^{7/12+\varepsilon}$ *where* $0 < \varepsilon < 5/12$ *is a fixed number. Then, for every* $M \in \mathbb{N}$,

$$\frac{1}{xy} \sum_{x \le n \le x+y} P(n) = \sum_{k=0}^{M} \frac{\xi_k}{\log^{k+1} x} + O\Big(\frac{1}{\log^{M+2} x}\Big),$$

where

$$\xi_k = \sum_{\nu=1}^{\infty} \frac{\log^k \nu}{\nu^2}. \tag{2.2}$$

THEOREM 6. *Let* $a \in \mathbb{Z} \setminus \{0\}$. *Then*

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \le x} e\big(\alpha s_q(P(n))\big) = 0.$$

Given an integer $n \ge 2$, write its prime factorisation as

$$n = P_r(n) P_{r-1}(n) \cdots P_1(n),$$

where $r = \Omega(n)$ (here $\Omega(n)$ stands for the number of prime factors of n counting multiplicity) and $P_r(n) \le P_{r-1}(n) \le \cdots \le P_1(n)$. Thus $P_j(n)$ is the jth largest prime factor of n, where for convenience $P_j(n) = 1$ if $j > \Omega(n)$.

THEOREM 7. *Let* $k \in \mathbb{N}$. *Let* $f_1(p), \dots, f_k(p)$ *be* k *functions defined on primes* p. *Assume that each* $f_i(p)$ *is bounded as* p *runs over* $\wp$, *and there exist constants* $C_1, \dots, C_k$ *for which*

$$S_j(x) := \sum_{p \le x} f_j(p) = (C_j + o(1)) \frac{x}{\log x} \qquad (x \to \infty).$$

Then

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \le x} \prod_{j=1}^{k} f_j(P_j(n)) = C_1 \cdots C_k.$$

3. Proof of Theorem 1. Let $N^{\lambda_N} \le H < 2H < N^{1-\lambda_N}$. If p is in $[H, 2H]$, then, by (iii),

$$\Big|\rho_N(p) - \frac{1}{\pi([H,2H])} \sum_{\substack{q \in [H,2H] \\ q \in \wp_N}} \rho_N(q)\Big| \le \varepsilon_N \rho_N(p),$$

where $\lim_{N\to\infty} \varepsilon_N = 0$. On the other hand, since the function is bounded, there exists an absolute constant $K > 0$ such that $|f(p)| \le K$ for all primes p.

Therefore,

$$\left|\sum_{\substack{p\in[H,2H]\\ p\in\wp_N}} f(p)\rho_N(p) - \frac{1}{\pi([H,2H])}\sum_{\substack{p,q\in[H,2H]\\ p,q\in\wp_N}} f(p)\rho_N(q)\right| \le 2K\varepsilon_N \sum_{p\in[H,2H]} \rho_N(p). \tag{3.1}$$

Moreover, using (2.1), we have

$$\begin{aligned}\frac{1}{\pi([H,2H])}&\sum_{\substack{p,q\in[H,2H]\\ p,q\in\wp_N}} f(p)\rho_N(q)\\ &= \sum_{\substack{q\in[H,2H]\\ q\in\wp_N}} \rho_N(q)\cdot\frac{1}{\pi([H,2H])}(S(2H)-S(H))\\ &= \left\{\eta\frac{\pi(2H)-\pi(H)}{\pi([H,2H])} + o\left(\frac{\pi(2H)}{\pi([H,2H])}\right)\right\}\sum_{\substack{q\in[H,2H]\\ q\in\wp_N}}\rho_N(q)\\ &= (\eta+o(1))\sum_{\substack{q\in[H,2H]\\ q\in\wp_N}}\rho_N(q) \quad \text{as } H,N\to\infty.\end{aligned} \tag{3.2}$$

Consider the sequence $H_0 = N^{\lambda_N}$, $H_{j+1} = 2H_j$ for each integer $0\le j\le J$ where J is such that $H_J \le N^{1-\lambda_N} \le 2H_J$. Then, in light of (3.1) and (3.2), as $N\to\infty$,

$$\begin{aligned}\sum_{\substack{p\in[H_0,H_J]\\ p\in\wp_N}} f(p)\rho_N(p) &= \sum_{j=0}^{J-1}\sum_{\substack{p\in[H_j,H_{j+1}]\\ p\in\wp_N}} f(p)\rho_N(p)\\ &= (\eta+o(1))\sum_{j=0}^{J-1}\sum_{\substack{q\in[H_j,H_{j+1}]\\ q\in\wp_N}}\rho_N(q)\\ &= (\eta+o(1))\sum_{\substack{q\in[H_0,H_J]\\ q\in\wp_N}}\rho_N(q).\end{aligned} \tag{3.3}$$

On the other hand, by conditions (i) and (ii) on $\rho_N(p)$,

$$\begin{aligned}\sum_{\substack{q\in[H_0,H_J]\\ q\in\wp_N}}\rho_N(q) &= 1 - \sum_{q<H_0}\rho_N(q) - \sum_{q>H_J}\rho_N(q)\\ &= 1 - o(1) - o(1) \quad \text{as } H,N\to\infty.\end{aligned} \tag{3.4}$$

Gathering (3.3) and (3.4) completes the proof of Theorem 1.

Remark 1. In the line of the function $f(x,p)$ defined in (1.4), let

$$\gamma_N(p) = \frac{1}{N}\#\{n \le N : P(n) = p\} = \frac{1}{N}\Psi\left(\frac{N}{p}, p\right),$$

where $\Psi(x,y) := \#\{n \le x : P(n) \le y\}$ for $2 \le y \le x$. Then one can easily check that $\gamma_N(p)$ is a prime weight function, since it satisfies (i)–(iv). More generally, given an integer $k \ge 1$ and recalling that $P_k(n)$ is the kth largest prime factor of n with $\Omega(n) \ge k$, we see that

$$\gamma_N^{(k)}(p) := \frac{1}{N}\#\{n \le N : P_k(n) = p\}$$

is also a prime weight function. This follows essentially by observing that

$$\gamma_N^{(k)}(p) = \frac{1}{N}\sum_{p_1 \ge \cdots \ge p_{k-1} \ge p} \Psi\left(\frac{N}{p_1 \cdots p_{k-1}}, p\right)$$

and then using the properties of $\Psi(x,y)$.

As consequences of Theorem 1, we have the following results.

Corollary 1. *Let k be a fixed positive integer and let $f : \wp \to \mathbb{C}$ be a bounded function satisfying* (2.1)*. Then, for some constant c_k,*

$$\frac{1}{N}\sum_{n \le N} f(P_k(n)) \to c_k \qquad (N \to \infty).$$

Corollary 2. *Let $(\varphi_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers for which the limit*

$$F(u) := \lim_{N\to\infty} \frac{1}{\pi(N)}\#\{p \in \wp_N : \varphi_p < u\}$$

exists, where $F(u)$ is a distribution function. Assume moreover that $\rho_N(p)$ is a prime weight function. Then

$$\lim_{N\to\infty} \sum_{\substack{p\in\wp_N \\ \varphi_p < u}} \rho_N(p) = F(u).$$

Proof. Indeed, one only needs to choose

$$f(p) = \begin{cases} 1 & \text{if } \varphi_p < u, \\ 0 & \text{otherwise,} \end{cases}$$

and then to apply Theorem 1. ∎

4. Proof of Theorem 2. Let $\rho_N(p)$ be a prime weight function and assume that if

$$F_N(u) := \sum_{\substack{p\in\wp_N \\ g(p+1) < u}} \rho_N(p),$$

then the limit

$$\lim_{N\to\infty} F_N(u) = F(u) \tag{4.1}$$

exists for almost all real numbers u and F is a distribution function. Then, since $F(-\infty) = 0$ and $F(\infty) = 1$, there exists a real number b for which the limit in (4.1) exists for $u = b$ and $u = b+1$ and satisfies $F(b+1) - F(b) > 0$. In this case, there exists a positive real number D such that

$$\lim_{N\to\infty} \sum_{\substack{p\in\wp_N \\ g(p+1)\in[b,b+1)}} \rho_N(p) = D.$$

It follows that there exists a sequence $(H_N)_{N\in\mathbb{N}}$ which tends to infinity with N and is such that $2H_N < N$ and, in light of condition (i),

$$\sum_{\substack{p\in[H_N,2H_N] \\ g(p+1)\in[b,b+1)}} \rho_N(p) > \frac{D}{2} \sum_{p\in[H_N,2H_N]} \rho_N(p),$$

thus implying that for some positive constant c, we have $Q_{\mathrm{pr}}(2H_N) > c$ for every positive integer N, where Q_{pr} is defined in (1.6). Hence $(W(2H_N))_{N\in\mathbb{N}}$ is a bounded sequence, where W is defined in (1.8). But this can only hold if $\lambda = 0$, in which case

$$\sum_p \frac{\min(1, g^2(p))}{p} < \infty. \tag{4.2}$$

Next, let

$$A_m := \sum_{\substack{p\le m \\ |g(p)|<1}} \frac{g(p)}{p} \qquad (m = 1, 2, \ldots). \tag{4.3}$$

It is known that (4.3) implies that $g(p+1) - A_p$ has a limiting distribution

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \le x : g(p+1) - A_p < u\} := L(u).$$

This implies that

$$\lim_{N\to\infty} \sum_{\substack{p\in\wp_N \\ g(p+1)-A_p<u}} \rho_N(p) = L(u).$$

In light of (4.2), we obtain

$$A_x - A_m = \sum_{\substack{m<p\le x \\ |g(p)|<1}} \frac{g(p)}{p}$$

and therefore

$$|A_x - A_m|^2 \le \sum_{m<p\le x} \frac{1}{p} \sum_{\substack{m<p\le x\\ |g(p)|<1}} \frac{g^2(p)}{p}.$$

Hence there exists λ_x which tends to 0 as $x \to \infty$ and if $m \ge x^{\lambda_x}$, then

$$|A_x - A_m|^2 \le \log\left(\frac{\log x}{\log x^{\lambda_x}}\right) \sum_{x^{\lambda_x} \le p \le x} \frac{g^2(p)}{p} \to 0 \quad \text{as } x \to \infty.$$

We will now prove that A_N is bounded as $N \to \infty$. Assume the contrary, that is, that there exists a sequence of positive integers $N_1 < N_2 < \cdots$ such that $A_{N_\nu} \to \infty$ as $\nu \to \infty$. Then, for every $\varepsilon > 0$,

$$\begin{aligned} L(u) &= \lim_{\nu\to\infty} \sum_{\substack{p\in\wp_{N_\nu}\\ g(p+1)<A_p+u}} \rho_{N_\nu}(p) \\ &\ge \lim_{\nu\to\infty} \sum_{\substack{p\in\wp_{N_\nu}\\ g(p+1)<A_{N_\nu}+u-\varepsilon}} \rho_{N_\nu}(p) - \sum_{p<N_\nu^{\lambda_\nu}} \rho_{N_\nu}(p) \\ &\ge \lim_{\nu\to\infty} \sum_{\substack{p\in\wp_{N_\nu}\\ g(p+1)<A_{N_\nu}+u-\varepsilon}} \rho_{N_\nu}(p) - \varepsilon \end{aligned} \tag{4.4}$$

since $\lambda_\nu \to 0$ as $\nu \to \infty$, where we have used condition (ii).

Now, since $A_{N_\nu} \to \infty$ as $\nu \to \infty$, given any large number E, we have $A_{N_\nu} \ge E$ provided ν is sufficiently large, in which case (4.4) yields

$$L(u) \ge F_{N_\nu}(E + u - \varepsilon) - \varepsilon,$$

implying that

$$L(u) \ge F(E + u - \varepsilon) - \varepsilon,$$

so that, since $\varepsilon > 0$ can be taken arbitrarily small,

$$L(u) \ge F(E + u). \tag{4.5}$$

As E can be chosen arbitrarily large, (4.5) yields $L(u) = 1$. Since this is true for every u, L cannot be a distribution function. The case $\liminf_{N\to\infty} A_N = -\infty$ can be treated similarly. We have thus established that $(A_N)_{N\in\mathbb{N}}$ is bounded.

We will now prove that $(A_N)_{N\in\mathbb{N}}$ is convergent. Indeed, suppose that

$$\limsup_{N\to\infty} A_N = \alpha \quad \text{and} \quad \liminf_{N\to\infty} A_N = \beta \quad \text{with } \alpha > \beta,$$

that is,

$$A_{M_\nu} \to \alpha \quad \text{and} \quad A_{N_\nu} \to \beta \quad \text{for two subsequences } A_{M_\nu} \text{ and } A_{N_\nu}.$$

We would then have

$$L(u) = \lim_{M_\nu\to\infty} \sum_{\substack{p\in\wp_{M_\nu}\\ g(p+1)<A_p+u}} \rho_{M_\nu}(p) \ge F(\alpha+u-\varepsilon)$$

and the above limit would also be $\le F(\alpha+u+\varepsilon)$, while

$$L(u) = \lim_{N_\nu\to\infty} \sum_{\substack{p\in\wp_{N_\nu}\\ g(p+1)<A_p+u}} \rho_{N_\nu}(p) \ge F(\beta+u-\varepsilon)$$

with the same limit $\le F(\beta+u+\varepsilon)$. This shows that we must have $\beta=\alpha$ and therefore

$$L(u) = F(\alpha+u).$$

Since A_m is bounded, we have proved that the series $\sum_{|g(p)|<1} g(p)/p$ is convergent, thus completing the proof of Theorem 2.

5. Proof of Theorem 3. Let $\lambda_x \to 0$ as $x\to\infty$ be a function to be chosen later, and set $T(x) := \sum_{n\le x} \tau(P(n)+a)$. We split this sum as follows:

$$\begin{aligned} T(x) &= \sum_{\substack{n\le x\\ P(n)\le x^{\lambda_x}}} \tau(P(n)+a) + \sum_{\substack{n\le x\\ x^{\lambda_x}<P(n)\le x^{1-\lambda_x}}} \tau(P(n)+a) \\ &\quad + \sum_{\substack{n\le x\\ x^{1-\lambda_x}<P(n)\le x}} \tau(P(n)+a) \\ &= S_1(x)+S_2(x)+S_3(x), \end{aligned}$$

say. Setting $M(x) := \sum_{p\le x} \tau(p+a)$ and using the estimate of $M(x)$ provided in (1.10), we get by partial summation

$$\begin{aligned} \sum_{p\le x} \frac{\tau(p+a)}{p} &= \int_{2-0}^{x} \frac{1}{u}\, dM(u) = \frac{M(u)}{u}\bigg|_{2-0}^{x} + \int_{2-0}^{x} \frac{M(u)}{u^2}\, du \\ &= \frac{D_a x + 2E_a x/\log x + O(x/\log^2 x)}{x} \\ &\quad + \int_{2-0}^{x} \left(\frac{D_a}{u} + \frac{2E_a}{u\log u} + O\left(\frac{1}{u\log^2 u}\right)\right) du \\ &= D_a + \frac{2E_a}{\log x} + O\left(\frac{1}{\log^2 x}\right) + D_a\log x + 2E_a\log\log x + O(1), \end{aligned}$$

from which it follows that

$$\sum_{p\le x} \frac{\tau(p+a)}{p} = D_a\log x + 2E_a\log\log x + O(1). \tag{5.1}$$

On the other hand, using the same technique, we get, for all $Y \geq 2$,

$$\sum_{Y \leq p < 2Y} \frac{\tau(p+a)}{p} = D_a \log 2 + O\left(\frac{1}{\log Y}\right). \tag{5.2}$$

We also easily establish that

$$\sum_{p \leq x} \frac{\tau(p+a)}{p \log p} = D_a \log\log x + O(1). \tag{5.3}$$

Using the well known estimate

$$\Psi(x,y) \leq cx \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\} \qquad (2 \leq y \leq x) \tag{5.4}$$

(see for instance De Koninck and Luca [5, Theorem 9.5]), we find that

$$\begin{aligned} S_1(x) = \sum_{p \leq x^{\lambda_x}} \tau(p+a)\Psi\left(\frac{x}{p}, p\right) &\leq cx \sum_{p \leq x^{\lambda_x}} \frac{\tau(p+a)}{p} \exp\left\{-\frac{1}{2}\frac{\log x}{\log p}\right\} \\ &\leq cx \exp\left\{-\frac{1}{2}\frac{1}{\lambda_x}\right\} \sum_{p \leq x^{\lambda_x}} \frac{\tau(p+a)}{p}, \end{aligned}$$

which combined with (5.1) and choosing

$$\lambda_x = \frac{1}{\log\log x} \tag{5.5}$$

yields

$$S_1(x) \ll \frac{x}{\sqrt{\log x}} \frac{1}{\log\log x} \log x = \frac{x\sqrt{\log x}}{\log\log x}. \tag{5.6}$$

On the other hand, using (5.1), we get

$$\begin{aligned} S_3(x) &= \sum_{x^{1-\lambda_x} < p \leq x} \tau(p+a)\Psi\left(\frac{x}{p}, p\right) \\ &\leq x \sum_{x^{1-\lambda_x} < p \leq x} \frac{\tau(p+a)}{p} \\ &= x(D_a \log x - D_a \log x^{1-\lambda_x} + O(\log\log x)) \ll \lambda_x x \log x. \end{aligned} \tag{5.7}$$

For the evaluation of $S_2(x)$, we proceed as follows. First, we set $J_x := (x^{\lambda_x}, x^{1-\lambda_x}]$. We may thus write

$$S_2(x) = \sum_{\substack{n \leq x \\ P(n) \in J_x}} \tau(P(n)+a) = \sum_{p \in J_x} \tau(p+a)\Psi\left(\frac{x}{p}, p\right). \tag{5.8}$$

Recalling the Hildebrand [14] estimate

$$\Psi(x,y) = x\rho(u)\left(1 + O\left(\frac{\log(u+1)}{\log y}\right)\right), \tag{5.9}$$

which is valid uniformly for $x \geq 3$, $\exp\{(\log\log x)^{5/3+\varepsilon}\} \leq y \leq x$, and setting $u_p = (\log x - \log p)/\log p$, we find that, for $p \in J_x$,

$$\Psi\left(\frac{x}{p}, p\right) = \frac{x}{p}\rho(u_p)\left(1 + O\left(\frac{\log(u_p+1)}{\log p}\right)\right) \tag{5.10}$$

and

$$u_p \in \left[\frac{\lambda_x}{1-\lambda_x}, \frac{1-\lambda_x}{\lambda_x}\right]. \tag{5.11}$$

Thus in light of (5.8), (5.10) and (5.11), we have

$$\begin{aligned} S_2(x) &= x\sum_{p\in J_x} \frac{\tau(p+a)}{p}\rho(u_p) + O\left(x\sum_{p\in J_x}\frac{\tau(p+a)}{p\log p}\log(u_p+1)\right) \\ &= x\sum_{p\in J_x} \frac{\tau(p+a)}{p}\rho(u_p) + O\left(x\log(1/\lambda_x)\sum_{p\in J_x}\frac{\tau(p+a)}{p\log p}\right) \\ &= xL(x) + O(xK(x)), \end{aligned} \tag{5.12}$$

say, where we have used the fact that $\log(u_p+1) \ll \log(1/\lambda_x)$ for $p \in J_x$.

On the other hand,

$$\begin{aligned} \sum_{n\leq x}\log P(n) &= \sum_{\substack{n\leq x\\ P(n)\leq x^{\lambda_x}}}\log P(n) + \sum_{\substack{n\leq x\\ x^{\lambda_x}<P(n)\leq x^{1-\lambda_x}}}\log P(n) \\ &\quad + \sum_{\substack{n\leq x\\ P(n)>x^{1-\lambda_x}}}\log P(n) \\ &= R_1(x) + R_2(x) + R_3(x), \end{aligned}$$

say. Since, using (5.4) and recalling our choice (5.5) of λ_x, we have

$$\begin{aligned} R_1(x) &= \sum_{p\leq x^{\lambda_x}}\log p\,\Psi\left(\frac{x}{p},p\right) \ll x\sum_{p\leq x^{\lambda_x}}\frac{\log p}{p}\exp\left\{-\frac{1}{2}\frac{\log x}{\log p}\right\} \\ &\ll x\exp\left\{-\frac{1}{2}\frac{1}{\lambda_x}\right\}\sum_{p\leq x^{\lambda_x}}\frac{\log p}{p} \ll \frac{x}{\sqrt{\log x}}\lambda_x\log x = x\frac{\sqrt{\log x}}{\log\log x} \end{aligned}$$

and similarly

$$R_3(x) \ll x\sum_{x^{1-\lambda_x}<p\leq x}\frac{\log p}{p} = (1+o(1))x\big(\log x - (1-\lambda_x)\log x\big) \ll \frac{x\log x}{\log\log x},$$

it follows that

$$\sum_{n\le x} \log P(n) = R_2(x) + o(x\log x),$$

which implies in light of (1.2) that

$$R_2(x) = \kappa x\log x + o(x\log x) \qquad (x\to\infty). \tag{5.13}$$

Since

$$R_2(x) = \sum_{p\in J_x} \log p\,\Psi\left(\frac{x}{p}, p\right) = x\sum_{p\in J_x} \frac{\log p}{p}\rho(u_p) + o(x\log x) \qquad (x\to\infty),$$

it follows from (5.13) that

$$\sum_{p\in J_x} \frac{\log p}{p}\rho(u_p) = \kappa\log x + o(\log x) \qquad (x\to\infty). \tag{5.14}$$

As a consequence of the Prime Number Theorem,

$$\sum_{Y\le p<2Y} \frac{\log p}{p} = \log 2 + O\left(\frac{1}{\log Y}\right).$$

Using this along with estimate (5.2) and the fact that

$$\max_{x^{\lambda_x}<p_1<p_2<2p_1<x^{1-\lambda_x}} \left|\frac{\rho(u_{p_1})}{\rho(u_{p_2})} - 1\right| \to 0 \qquad \text{as } x\to\infty,$$

we obtain

$$\left|\sum_{Y\le p<2Y} \frac{\tau(p+a)}{p}\rho(u_p) - D_a\sum_{Y\le p<2Y} \frac{\log p}{p}\rho(u_p)\right| = O\left(\frac{\rho(u_Y)}{\log Y}\right) \qquad (Y\ge 2). \tag{5.15}$$

Let us now define

$$H_0 = x^{\lambda_x}, \qquad H_j = 2^j H_0 \quad \text{for } j = 1,\dots,\mathcal{I},$$

where $H_{\mathcal{I}-1} < x^{1-\lambda_x} \le H_{\mathcal{I}}$, so that $\mathcal{I} = \lceil \frac{(1-2\lambda_x)\log x}{\log 2} \rceil$. Hence, (5.15) yields

$$\left|\sum_{H_0\le p<H_{\mathcal{I}}} \frac{\tau(p+a)}{p}\rho(u_p) - D_a\sum_{H_0\le p<H_{\mathcal{I}}} \frac{\log p}{p}\rho(u_p)\right| \le \sum_{j=0}^{\mathcal{I}-1} O\left(\frac{\rho(u_j)}{\log H_j}\right) \ll \frac{\mathcal{I}}{\log H_0} \ll \frac{1}{\lambda_x}. \tag{5.16}$$

Since clearly

$$\left|\sum_{x^{1-\lambda_x}<p<H_{\mathcal{I}}} \left(\frac{\tau(p+a)}{p} - D_a\frac{\log p}{p}\right)\rho(u_p)\right| \ll 1,$$

it follows from (5.16) that

$$\left|\sum_{p\in J_x}\frac{\tau(p+a)}{p}\rho(u_p)-D_a\sum_{p\in J_x}\frac{\log p}{p}\rho(u_p)\right|\ll\frac{1}{\lambda_x}. \tag{5.17}$$

Using (5.17) and (5.14), we get

$$L(x)=\kappa D_a\log x+O(1/\lambda_x). \tag{5.18}$$

On the other hand, using (5.3), we deduce that

$$K(x)\ll\log(1/\lambda_x)\sum_{p\in J_x}\frac{\tau(p+a)}{p\log p}\ll\log(1/\lambda_x)\log\log x. \tag{5.19}$$

Combining (5.18) and (5.19) in (5.12) yields

$$S_2(x)=\kappa D_a x\log x+o(x\log x)\qquad(x\to\infty). \tag{5.20}$$

Gathering estimates (5.6), (5.7) and (5.20) completes the proof of Theorem 3.

6. Proof of Theorem 4. The proof of Theorem 4 is similar to that of Theorem 3 and we will therefore omit it.

7. Proof of Theorem 5. It is clear that in order to prove our result, we may assume that $y=x^\lambda$, with $7/12<\lambda<11/12$, say.

It follows from Corollary 1 in Ramachandra, Sankaranarayanan and Srinivas [22] that

$$\sum_{n<x\le x+y}\Lambda(n)=y+O(y\exp\{-(\log x)^{1/6}\}). \tag{7.1}$$

Now, observe that if $p\in[x,x+y]$, then

$$\log x\le\log p\le\log x+\log(1+y/x)=\log x+O(y/x)$$

while

$$\begin{aligned}\sum_{\substack{x\le p^\ell<x+y\\ \ell\ge2}}\log p&\le(2\log x)(\sqrt{x+y}-\sqrt{x})+O(x^{1/3})\\&\le\frac{2(\log x)y}{\sqrt{x}}+O(x^{1/3})\ll x^{1/3},\end{aligned}$$

so that, using (7.1), we get

$$\sum_{x\le p\le x+y}(\log x+O(y/x))+O((y/x)(\log x))=y+O(y\exp\{-(\log x)^{1/6}\}),$$

which then allows us to write

$$\sum_{p\in[x,x+y]} 1 = \frac{y}{\log x} + O\biggl(\frac{y}{x}\sum_{p\in[x,x+y]} 1\biggr) + O(y\exp\{-(\log x)^{1/6}\})$$
$$= \frac{y}{\log x} + O\biggl(\frac{y}{x}\,\frac{y}{\log x}\biggr) + O(y\exp\{-(\log x)^{1/6}\}).$$

Consequently,

$$(7.2)\qquad \sum_{p\in[x,x+y]} p = \frac{xy}{\log x} + O\biggl(\frac{y^2}{\log x}\biggr) + O(xy\exp\{-(\log x)^{1/6}\})$$
$$= \frac{xy}{\log x} + O(xy\exp\{-(\log x)^{1/6}\}),$$

where we have used the fact that $y^2/\log x < xy\exp\{-(\log x)^{1/6}\}$.

Now, provided that $7/12 + \varepsilon_1 < \log v/\log u < 11/12$, say, where $\varepsilon_1 > 0$ is an arbitrarily small number, we have

$$(7.3)\qquad \sum_{x\le n\le x+y} P(n) = \sum_{\substack{x\le \nu p\le x+y\\ P(\nu)\le p}} p$$
$$= \sum_{\nu<x^{\varepsilon_2}}\ \sum_{x/\nu<p\le x/\nu+y/\nu} p + \sum_{x^{\varepsilon_2}\le\nu\le x}\ \sum_{x/\nu<p\le x/\nu+y/\nu} p$$
$$= S_1(x,y) + S_2(x,y),$$

say. Now it is clear that

$$(7.4)\qquad S_2(x,y) \le x^{1-\varepsilon_2}y.$$

On the other hand, writing

$$(7.5)\qquad S_1(x,y) = \sum_{\nu<x^{\varepsilon_2}} A_\nu,$$

say, and assuming that $\frac{\log x}{\log y} < \frac{\log(y/\nu)}{\log(x/\nu)} < \frac{\log x}{\log y} + \varepsilon$ (which holds if in (7.5), ν runs from 1 to x^{ε_2} for some positive ε_2 sufficiently small), we deduce from (7.2) that

$$(7.6)\qquad A_\nu = \frac{xy}{\nu^2\log(x/\nu)} + O\biggl(\frac{xy}{\nu^2}\exp\{-(\log\sqrt{x})^{1/6}\}\biggr),$$

where we have used the fact that $\log(x/\nu) > \log\sqrt{x}$.

It follows from (7.5) and (7.6) that, for some positive constant c,

$$(7.7)\qquad S_1(x,y) = \sum_{\nu<x^{\varepsilon_2}} \frac{xy}{\nu^2\log(x/\nu)} + O(xy\exp\{-(c\log x)^{1/6}\}).$$

Now, observe that

$$
\begin{aligned}
(7.8)\qquad T &:= \sum_{\nu<x^{\varepsilon_2}} \frac{1}{\nu^2 \log(x/\nu)} = \sum_{\nu<x^{\varepsilon_2}} \frac{1}{\nu^2 \log x} \frac{1}{1-\frac{\log \nu}{\log x}} \\
&= \sum_{k=0}^{\infty} \sum_{\nu<x^{\varepsilon_2}} \frac{1}{\nu^2 \log x} \frac{\log^k \nu}{\log^k x} \\
&= \sum_{k=0}^{M} \frac{1}{\log^{k+1} x} \sum_{\nu<x^{\varepsilon_2}} \frac{\log^k \nu}{\nu^2} + O\Big(\sum_{\nu} \frac{\log^{M+1} \nu}{\nu^2 \log^{M+2} x}\Big).
\end{aligned}
$$

We easily see that, for each integer $k \geq 0$, by partial integration,

$$
\begin{aligned}
(7.9)\qquad J_k(z) &:= \int_z^{\infty} \eta^k e^{-\eta}\, d\eta = \eta^k(-e^{-\eta})|_z^{\infty} + k\int_z^{\infty} \eta^{k-1} e^{-\eta}\, d\eta \\
&= z^k e^{-z} + kJ_{k-1}(z)
\end{aligned}
$$

with, in particular, $J_0(z) = e^{-z}$.

Setting

$$
R_k := \sum_{\nu \geq x^{\varepsilon_2}} \frac{\log^k \nu}{\nu^2}
$$

and using (7.9) clearly shows that

$$
(7.10)\qquad R_k \leq 2\int_{x^{\varepsilon_2}}^{\infty} \frac{\log^k t}{t^2}\, dt = 2\int_{\varepsilon_2 \log x}^{\infty} \eta^k e^{-\eta}\, d\eta = 2J_k(\varepsilon_2 \log x).
$$

Assuming that M is fixed, we deduce from (7.10) that

$$
(7.11)\qquad R_k \ll (\varepsilon_2 \log x)^k e^{-\varepsilon_2 \log x} \qquad (k \leq M).
$$

Recalling the definition of ξ_k given in (2.2), and using (7.11) in (7.8), it follows that

$$
(7.12)\qquad T = \sum_{k=0}^{M} \frac{\xi_k}{\log^{k+1} x} + O\Big(\frac{1}{\log^{M+2} x}\Big).
$$

Using (7.12) in (7.7), and substituting the resulting estimate in (7.3), taking into account estimate (7.4), completes the proof of Theorem 5.

8. Proof of Theorem 6. Setting $f(p) := e(\alpha s_q(p))$, it has been shown by Mauduit and Rivat [19] that

$$
\lim_{x\to\infty} \frac{1}{\pi(x)} \sum_{p \leq x} f(p) = 0.
$$

From this and Theorem 1, Theorem 6 follows.

9. Proof of Theorem 7. Given $J \subseteq \wp$, we set $\omega_J(n) := \#\{p \in J : p \mid n\}$. Let δ be a small positive number. Given a large number x and setting $J = J_x = [x^\delta, x]$, it follows from the Turán–Kubilius inequality that

$$\sum_{n \le x} \Bigl(\omega_J(n) - \sum_{p \in J} \frac{1}{p} \Bigr)^2 \le Cx \sum_{x^\delta \le p \le x} \frac{1}{p}.$$

Since

$$\sum_{n \le x} \sum_{\substack{p \mid n \\ x^{1-\delta} \le p \le x}} 1 \le x \sum_{x^{1-\delta} \le p \le x} \frac{1}{p} \le x \Bigl(\log \frac{1}{1-\delta} + o(1) \Bigr) \le 2\delta x,$$

provided x is large enough, and since

$$\sum_{x^\delta \le p \le x} \frac{1}{p} = \log \frac{1}{\delta} + o(1) \qquad (x \to \infty),$$

it follows that there exists an absolute constant $c > 0$ and a number x_0 such that if $x > x_0$, then

$$x^\delta \le P_k(n) < \cdots < P_1(n) \le x^{1-\delta}$$

for every integer $n \in [2, x]$ with the exception of at most $c\delta x$ integers.

Now let λ be a small positive number such that $\lambda \log 1/\delta \le \delta$, and let us consider the set D_x of those positive integers $n \le x$ which have two prime divisors p, q such that $x^\delta < p < q < p^{1+\lambda}$. It turns out that

$$\begin{aligned} \#D_x &\le x \sum_{x^\delta < p < q < p^{1+\lambda}} \frac{1}{pq} \le x \sum_{x^\delta < p \le \sqrt{x}} \frac{1}{p} \sum_{p < q < p^{1+\lambda}} \frac{1}{q} \\ &\le 2x \sum_{x^\delta < p \le x} \frac{\log(1+\lambda)}{p} \le 2x\lambda \log \frac{1}{\delta} \le 2\delta x. \end{aligned}$$

Let $\mathcal{B} = \mathcal{B}_x$ be the set of those k-tuples of primes $(p_1, \ldots, p_k)$ such that

$$x^\delta \le p_k < \cdots < p_1 \le x^{1-\delta}, \quad p_{j+1} < p_j^{1-\lambda} \text{ for } j = 1, \ldots, k-1, \quad p_1 \cdots p_k < x^{1-\delta}.$$

First observe that the size of the set of those positive integers $n \le x$ for which the k-tuples $(P_1(n), \ldots, P_k(n))$ are not in $\mathcal{B}$ is $O(\delta x)$. Thus

$$\begin{aligned} T &:= \sum_{n \le x} \prod_{j=1}^{k} f_j(P_j(n)) \\ &= \sum_{(p_1, \ldots, p_k) \in \mathcal{B}} \prod_{j=1}^{k} f_j(p_j) \Psi\Bigl(\frac{x}{p_1 \cdots p_k}, p_k \Bigr) + O(\delta x). \end{aligned} \tag{9.1}$$

Then, if $(p_1, \ldots, p_k) \in \mathcal{B}$, one can easily see that with

$$u = \frac{\log(x/p_1 \cdots p_k)}{\log p_k},$$

we get

$$u = \frac{\log x - \sum_{j=1}^{k} \log p_j}{\log p_k} \le \frac{1}{\delta},$$

so that

$$\frac{\log(u+1)}{\log y} = \frac{\log(u+1)}{\log p_k} \le \frac{\log(1/\delta)}{\delta \log x}.$$

We thus obtain, using (5.9),

$$\begin{aligned}\Psi\Big(\frac{x}{p_1\cdots p_k}, p_k\Big) \\ = \frac{x}{p_1\cdots p_k}\rho\Big(\frac{\log(x/p_1\cdots p_k)}{\log p_k}\Big) + O\Big(\frac{x}{p_1\cdots p_k}\,\frac{\log(1/\delta)}{\delta}\,\frac{1}{\log x}\Big).\end{aligned}$$

Hence, it follows from (9.1) that

$$\begin{aligned}T = \sum_{(p_1,\ldots,p_k)\in\mathcal{B}} \prod_{j=1}^{k} f_j(p_j)\frac{x}{p_1\cdots p_k}\rho\Big(\frac{\log(x/p_1\cdots p_k)}{\log p_k}\Big) \\ + O\Big(\sum_{(p_1,\ldots,p_k)\in\mathcal{B}} \prod_{j=1}^{k} f_j(p_j)\frac{x}{p_1\cdots p_k}\,\frac{\log(1/\delta)}{\delta\log x}\Big).\end{aligned} \tag{9.2}$$

Since the above error term is, as $x \to \infty$,

$$\begin{aligned}&\ll \frac{\log(1/\delta)}{\delta\log x} \sum_{(p_1,\ldots,p_k)\in\mathcal{B}} \prod_{j=1}^{k} f_j(p_j)\Big(\sum_{x^\delta<p_j<x^{1-\delta}} \frac{1}{p_j}\Big)^k \\ &\ll \frac{\log(1/\delta)}{\delta\log x}\Big(\log\frac{1-\delta}{\delta}\Big)^k = o(x),\end{aligned}$$

it follows, in light of (9.2), that estimate (9.1) can be replaced by

$$\begin{aligned}T = x \sum_{(p_1,\ldots,p_k)\in\mathcal{B}} \prod_{j=1}^{k} \frac{f_j(p_j)}{p_j}\rho\Big(\frac{\log(x/p_1\cdots p_k)}{\log p_k}\Big) \\ + O(\delta x) + o(x) \quad (x\to\infty).\end{aligned} \tag{9.3}$$

Now, given any k primes $q_1 < \cdots < q_k$ with $1/2 < q_j/p_j < 2$ for $j = 1,\ldots,k$ and setting

$$\begin{aligned}\varepsilon(x) := \max_{(p_1,\ldots,p_k)\in\mathcal{B}} \max_{\substack{q_1,\ldots,q_k \\ q_j/p_j\in(1/2,2)}} \Big|\rho\Big(\frac{\log x - \sum_{j=1}^{k}\log q_j}{\log q_k}\Big) \\ - \rho\Big(\frac{\log x - \sum_{j=1}^{k}\log p_j}{\log p_k}\Big)\Big|,\end{aligned}$$

it follows from the continuity of ρ that $\varepsilon(x) \to 0$ as $x \to \infty$. We can then use this in the estimate of the main term in (9.3) so that, arguing as in the proof of Theorem 1, we deduce that (9.3) can be replaced by

$$\begin{aligned} T &= C_1 \cdots C_k \sum_{(p_1,\ldots,p_k)\in\mathcal{B}} x \prod_{j=1}^{k} \frac{1}{p_j} \rho\left(\frac{\log(x/p_1\cdots p_k)}{\log p_k}\right) + O(\delta x) + o(x) \\ &= C_1 \cdots C_k \left(x + O(\delta x)\right) + O(\delta x) + o(x) = C_1 \cdots C_k x + O(\delta x). \end{aligned}$$

Since δ can be chosen arbitrarily small, the proof of Theorem 7 is complete.

10. Final remarks. Given a real-valued additive function g and $a \in \mathbb{Z} \setminus \{0\}$, let

$$F_N(y) := \frac{1}{\pi(N)} \#\{p \in \mathbb{N} : g(p+a) < y\} \quad \text{and} \quad F(y) = \lim_{N\to\infty} F_N(y).$$

We then have the following results.

THEOREM 8. *Given arbitrary real numbers $y_1, \ldots, y_k$,*

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N : g(P_j(n)+a) < y_j,\ j = 1,\ldots,k\} = \prod_{j=1}^{k} F(y_j).$$

THEOREM 9. *Given any real number z, set $G(z) := \int_{-\infty}^{\infty} F(y+z)\, dF(y)$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N : g(P_1(n)+a) - g(P_2(n)+a) < z\} = G(z).$$

THEOREM 10. *Let $a_1, \ldots, a_k$ be non-zero integers and let $g_1, \ldots, g_k$ be real-valued additive functions each satisfying the three-series condition* (1.5). *Set*

$$F_{N,j}(y) := \frac{1}{\pi(N)} \#\{p \le N : g_j(p+a_j) < y\}.$$

Then, for each $j \in \{1, \ldots, k\}$, we have $\lim_{N\to\infty} F_{N,j}(y) = F_j(y)$. Moreover, given any real numbers $y_1, \ldots, y_k$, the limit

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N : g_j(P_j(n)+a_j) < y_j,\ j = 1,\ldots,k\}$$

exists and is equal to $\prod_{j=1}^{k} F_j(y_j)$.

THEOREM 11. *Let $a_1, \ldots, a_k$ be non-zero integers and let $g_1, \ldots, g_k$ be real-valued additive functions each satisfying $g_j(p) = O(1)$ for $p \in \wp$. Let*

$$A_j(x) := \sum_{p\le x} \frac{g_j(p)}{p} \quad \text{and} \quad B_j(x)^2 := \sum_{p\le x} \frac{g_j^2(p)}{p}$$

and assume $B_j(x)\to\infty$ as $x\to\infty$. Then, for any real numbers $y_1,\dots,y_k$,

$$\lim_{N\to\infty}\frac{1}{N}\#\left\{n\le N:\frac{g_j(P_j(n)+a_j)-A_j(N)}{B_j(N)}<y_j,\ j=1,\dots,k\right\}=\prod_{j=1}^{k}\Phi(y_j).$$

The above theorems are essentially consequences of Theorem 7. For instance, to prove Theorem 10, one can proceed as follows. First define

$$f_j(p)=\begin{cases}1 & \text{if } g_j(p+a_j)<y_j,\\ 0 & \text{otherwise}\end{cases}\qquad (j=1,\dots,k).$$

Then

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\sum_{p\le N}f_j(p)=F_j(y_j)\qquad (j=1,\dots,k),$$

implying that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n\le N}f_j(P_j(n))=F_j(y_j)\qquad (j=1,\dots,k).$$

It follows that, using Theorem 7, we get

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n\le N}\prod_{j=1}^{k}f_j(P_j(n))=\prod_{j=1}^{k}F_j(y_j),$$

which is precisely the conclusion of Theorem 10.

To prove Theorem 9, we first observe that $g(P_1(n)+a)$ and $g(P_2(n)+a)$ are independent. Then, applying the result of Theorem 8, the conclusion of Theorem 9 follows.

The proof of Theorem 11 requires more attention. First we let

$$h_j(p)=\frac{g_j(p+a_j)-A_j(p)}{B_j(p)}$$

and define

$$f_j(p)=\begin{cases}1 & \text{if } h_j(p)<y_j,\\ 0 & \text{otherwise}\end{cases}\qquad (j=1,\dots,k).$$

Now, it is known that

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\#\left\{p\le N:\frac{g_j(p+a_j)-A_j(N)}{B_j(N)}<y_j\right\}=\Phi(y_j)\qquad (j=1,\dots,k).$$

On the other hand, it is clear that, for every $\varepsilon>0$,

$$\lim_{N\to\infty}\frac{1}{\pi(N)}\#\left\{p\le N:\left|\frac{g_j(p+a_j)-A_j(N)}{B_j(N)}-\frac{g_j(p+a_j)-A_j(p)}{B_j(p)}\right|>\varepsilon\right\}=0.$$

From this,

$$\lim_{N\to\infty} \frac{1}{\pi(N)} \sum_{p\le x} f_j(p) = \Phi(y_j) \qquad (j = 1, \dots, k).$$

On the other hand, it is a consequence of Theorem 8 that, for every $\varepsilon > 0$,

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ n \le N : \max_{j=1,\dots,k} \left| \frac{g_j(P_j(n)+a_j) - A_j(P_j(n))}{B_j(P_j(n))} - \frac{g_j(P_j(n)+a_j) - A_j(N)}{B_j(N)} \right| > \varepsilon \right\} = 0.$$

Combining the above estimates completes the proof of Theorem 11.

Acknowledgements. The authors would like to thank the referee for some valuable comments and corrections which helped improve the quality of this paper.

The work of the first author was partially supported by a grant from NSERC.

References

[1] E. Bombieri, J. B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. 156 (1986), 203–251.

[2] B. M. Bredikhin, *The dispersion method and binary additive problems of definite type*, Uspekhi Mat. Nauk 20 (1965), no. 2 (122), 89–130 (in Russian).

[3] J.-M. De Koninck, *On the largest prime divisors of an integer*, in: Extreme Value Theory and Applications (Gaithersburg, MD, 1993), Kluwer, Dordrecht, 1994, 447–462.

[4] J.-M. De Koninck and A. Ivić, *The distribution of the average prime divisor of an integer*, Arch. Math. (Basel) 43 (1984), 37–43.

[5] J.-M. De Koninck and F. Luca, *Analytic Number Theory: Exploring the Anatomy of Integers*, Grad. Stud. Math. 134, Amer. Math. Soc., Providence, RI, 2012.

[6] J.-M. De Koninck and R. Sitaramachandrarao, *Sums involving the largest prime divisor of an integer*, Acta Arith. 48 (1987), 1–8.

[7] J.-M. De Koninck and J. Sweeney, *On the unimodal character of the frequency function of the largest prime factor*, Colloq. Math. 88 (2001), 159–174.

[8] P. D. T. A. Elliott, *Probabilistic Number Theory I: Mean-Value Theorems*, Grundlehren Math. Wiss. 239, Springer, New York, 1979.

[9] P. D. T. A. Elliott, *Probabilistic Number Theory II: Central Limit Theorems*, Grundlehren Math. Wiss. 240, Springer, New York, 1980.

[10] P. D. T. A. Elliott, *The concentration function of additive functions on shifted primes*, Acta Math. 173 (1994), 1–35.

[11] P. Erdős, A. Ivić, and C. Pomerance, *On sums involving reciprocals of the largest prime factor of an integer*, Glas. Mat. (3) 21 (1986), 283–300.

[12] É. Fouvry, *Sur le problème des diviseurs de Titchmarsh*, J. Reine Angew. Math. 357 (1985), 51–76.

[13] A. Hildebrand, *Additive and multiplicative functions on shifted primes*, Proc. London Math. Soc. (3) 59 (1989), 209–232.
[14] A. Hildebrand, *On the number of positive integers $\leq x$ and free of prime factors $> y$*, J. Number Theory 22 (1986), 289–307.
[15] C. Hooley, *On the representation of a number as the sum of two squares and a prime*, Acta Math. 97 (1957), 189–210.
[16] I. Kátai, *On distribution of arithmetical functions on the set prime plus one*, Compos. Math. 19 (1968), 278–289.
[17] I. Kátai, *Some remarks on additive arithmetical functions*, Litovsk. Mat. Sb. 9 (1969), 515–518.
[18] Yu. V. Linnik, *The Dispersion Method in Binary Additive Problems*, Amer. Math. Soc., Providence, RI, 1963.
[19] C. Mauduit et J. Rivat, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. of Math. (2) 171 (2010), 1591–1646.
[20] N. McNew, *Popular values of the largest prime divisor function*, arXiv:1504.05985 (2015).
[21] E. Naslund, *The average largest prime factor*, Integers 13 (2013), Paper No. A81, 5 pp.
[22] K. Ramachandra, A. Sankaranarayanan, and K. Srinivas, *Addendum to Ramachandra's paper "Some problems of analytic number theory"*, Acta Arith. 73 (1995), 367–371.
[23] I. Z. Ruzsa, *On the concentration of additive functions*, Acta Math. Acad. Sci. Hungar. 36 (1980), 215–232.
[24] N. M. Timofeev, *The Erdős–Kubilius conjecture on the distribution of the values of additive functions on sequences of shifted primes*, Acta Arith. 58 (1991), 113–131 (in Russian).

Jean-Marie De Koninck
Département de mathématiques et de statistique
Université Laval
Québec G1V 0A6, Canada
E-mail: jmdk@mat.ulaval.ca

Imre Kátai
Computer Algebra Department
Eötvös Loránd University
Pázmány Péter Sétány I/C
1117 Budapest, Hungary
E-mail: katai@inf.elte.hu