On the counting function of semiprimes

Dragoș Crișan and Radek Erban

Abstract

A semiprime is a natural number which can be written as the product of two primes. The asymptotic behaviour of the function $\pi_2(x)$, the number of semiprimes less than or equal to x, is studied. Using a combinatorial argument, asymptotic series of $\pi_2(x)$ is determined, with all the terms explicitly given. An algorithm for the calculation of the constants involved in the asymptotic series is presented and the constants are computed to 20 significant digits. The errors of the partial sums of the asymptotic series are investigated. A generalization of this approach to products of k primes, for $k \geq 3$, is also proposed.

1. *Introduction*

For a positive integer k and a positive integer (or real number) x, let $\pi_k(x)$ be the number of integers less than or equal x which can be written as the product of k prime factors. The behaviour of $\pi_k(x)$ has been extensively studied during last two centuries, with the main focus on the case $k = 1$, where $\pi_1(x)$ is the prime counting function, denoted $\pi(x)$ in the rest of this paper. The prime number theorem states that $\pi(x) \sim \mathrm{li}(x)$, where the logarithmic integral function $\mathrm{li}(x) = \int_0^x \log^{-1} t \,\mathrm{d}t$ can be written as an asymptotic expansion $\mathrm{li}(x) \sim \frac{x}{\log x} \sum_{n=0}^{\infty} \frac{n!}{(\log x)^n}$. Bounds on the error term have been established in the literature, including the recent work of Trudgian [**22**], who proved that, for sufficiently large x,

$$\big|\pi(x) - \mathrm{li}(x)\big| \;\leq\; 0.2795\,\frac{x}{(\log x)^{3/4}}\;\exp\left(-\sqrt{\frac{\log x}{6.455}}\right).$$

This implies the existence of constants d_1 and d_2 such that

$$\big|\pi(x) - \mathrm{li}(x)\big| \;\leq\; d_1\,\frac{x}{(\log x)^{3/4}}\;\exp\left(-d_2\sqrt{\log x}\right), \qquad \text{for all } x \geq 2. \tag{1.1}$$

Assuming the Riemann hypothesis, Rosser and Schoenfeld [**19**, **20**] established even sharper bounds on the error term, including

$$|\pi(x) - \mathrm{li}(x)| < \frac{\sqrt{x}\log x}{8\pi}$$

for large enough x. Other explicit estimates of $\pi(x)$, in terms of x and $\log x$ are achievable, as proved by Axler [**1**].

In this paper, we focus on the case $k = 2$, where the numbers written as products of two (not necessarily distinct) primes are called semiprimes. In this case, Ishmukhametov and Sharifullina [**14**] recently used probabilistic arguments to approximate the behaviour of $\pi_2(x)$ as

$$\pi_2(x) \approx \frac{x\log(\log x)}{\log x} + 0.265\,\frac{x}{\log x} - 1.540\,\frac{x}{(\log x)^2}\,. \tag{1.2}$$

2000 *Mathematics Subject Classification* 11N37 (primary), 11K65, 11Y60 (secondary). .

The first term of (1.2) has already been known to Landau [**15**, §56], with his result stated, for general $k \in \mathbb{N}$, as

$$\pi_k(x) \sim \frac{1}{(k-1)!} \frac{x\,(\log(\log x))^{k-1}}{\log x}\,. \tag{1.3}$$

Delange [**6**, Theorem 1] obtained the asymptotic expansion of $\pi_k(x)$ in the form

$$\pi_k(x) \sim \frac{x}{\log x} \sum_{n=0}^{\infty} \frac{P_{n,k}(\log(\log x))}{(\log x)^n}\,, \tag{1.4}$$

where $P_{n,k}$ are polynomials of degree $k-1$, with the leading coefficient equal to $n!/(k-1)!$. Tenenbaum [**21**] proved a similar result, giving an expression for the coefficients in the polynomial $P_{0,k}$ in terms of the derivatives of $\frac{1}{\Gamma(z+1)} \prod_p \left(1 + \frac{z}{p-1}\right)\left(1-\frac{1}{p}\right)^z$ evaluated at $z = 0$. Considering $k = 2$ in (1.4), we can write an asymptotic expansion for $\pi_2(x)$ as

$$\pi_2(x) \sim \sum_{n=1}^{\infty} (n-1)!\, \frac{x \log(\log x)}{(\log x)^n} + \sum_{n=1}^{\infty} C_{n-1}\, \frac{x}{(\log x)^n}\,. \tag{1.5}$$

In Theorem 2.3, we prove that $C_0 = M$, where $M = 0.261497...$ is the Meissel–Mertens constant defined by

$$M = \lim_{x\to\infty} \left(\sum_{p \le x} \frac{1}{p} - \log(\log x) \right), \tag{1.6}$$

where we sum over all primes such that $p \le x$. In Section 3, we calculate the rest of constants C_n appearing in equation (1.5). They are given in Table 1 and obtained by the formula

$$C_n = n! \left(\sum_{i=0}^{n} \frac{B_i}{i!} - \sum_{i=1}^{n} \frac{1}{i} \right) = n! \left(\sum_{i=0}^{n} \frac{B_i}{i!} - H_i \right), \tag{1.7}$$

where H_i is the i-th harmonic number, $B_0 = M$ and constants B_i are defined using the asymptotic behaviour of sums [**15**, §56]

$$\sum_{p \le x} \frac{(\log p)^i}{p} = \frac{(\log x)^i}{i} + B_i + \mathcal{O}\Big(e^{-\sqrt[14]{\log x}}\Big)\,, \qquad \text{for } i \in \mathbb{N}. \tag{1.8}$$

Constants B_i are given as limits (3.1) in Section 3, where we present an algorithm to efficiently calculate them to a desired accuracy. They are computed in Table 1 to 20 significant digits. Rosser and Schoenfeld [**18**] prove that the error term in (1.8) can be given explicitly in terms of an integral, which contains the error terms in the prime number theorem. For the case $i = 1$ in equation (1.8), explicit estimates of this sum and, in particular, of the constant B_1 involved, were recently obtained by Dusart [**8**].

A related arithmetic function, $\Omega(m)$, is defined to be the number of prime divisors of $m \in \mathbb{N}$, where prime divisors are counted with their multiplicity. Considering fixed x in equation (1.3), we can view this approximation of $\pi_k(x)/x$ as the probability mass function of the Poisson distribution with mean $\log(\log(x))$. Erdős and Kac [**9**] showed that the distribution of $\Omega(x)$ is Gaussian with mean $\log(\log(x))$ (see also Rényi and Turán [**17**] and Harper [**13**] for generalizations and better bounds). Diaconis [**7**] obtained the asymptotic expansions for the average number of prime divisors as (see also Finch [**11**, Section 1.4.3])

$$\frac{1}{x} \sum_{m \le x} \Omega(m) \;\sim\; \log(\log x) + 1.0346538818\cdots + \sum_{n=1}^{\infty} \left(-1 + \sum_{i=0}^{n-1} \frac{\gamma_i}{i!} \right) \frac{(n-1)!}{\log^n x}\,, \tag{1.9}$$

where the constants γ_i are the Stieltjes constants, numerically computed in [**3**] to 20 significant digits. An asymptotic series for the variance of Ω have also been obtained [**11**, Section 1.4.3].

The Stieltjes constants γ_i are used in Section 3 during our calculation of the values of constants B_n and C_n, for $n \in \mathbb{N}$.

This paper is organized as follows. Section 2 begins with a counting lemma for expressing the semiprime counting function π_2 in terms of the prime counting function π. Using this lemma, the main results on the asymptotic behaviour of π_2 are stated and proved in Section 2 as Theorem 2.3 and Theorem 2.5. While Theorem 2.3 only gives the first two terms, its proof is more coincise than the proof of Theorem 2.5, which gives the full asymptotic series of π_2. The constants C_n which appear in this asymptotic series are computed in Section 3, where we present an efficient approach to calculate both constants B_n and C_n, based on the differentiation of the prime zeta function. In Section 4, we investigate the behaviour of the error terms given by the partial sums of the asymptotic series of π_2. We conclude with a generalization of the counting argument in Section 5, discussing the extensions of the presented results to the general case of counting functions π_k for $k \geq 3$.

2. *Asymptotic behaviour of the counting function of semiprimes*

As in equation (1.6), we denote primes by p and the sums over p shall be understood as sums over all primes satisfying the given condition. In the case of summing over primes twice, we denote the corresponding prime summation indices by p_1 and p_2. We begin with a simple counting formula [**14**], that gives a way of computing $\pi_2(x)$.

LEMMA 2.1. *For a positive integer x, the following holds*

$$\pi_2(x) = \frac{\pi(\sqrt{x}) - \pi(\sqrt{x})^2}{2} + \sum_{p \leq \sqrt{x}} \pi\left(\frac{x}{p}\right). \tag{2.1}$$

Proof. By the definition of counting functions π_2 and π, we have

$$\pi_2(x) = \sum_{\substack{p_1 \leq p_2 \\ p_1 p_2 \leq x}} 1 = \sum_{p_1 \leq \sqrt{x}} \sum_{p_1 \leq p_2 \leq \frac{x}{p_1}} 1 = \sum_{p_1 \leq \sqrt{x}} \left(\pi\left(\frac{x}{p_1}\right) - \pi(p_1) + 1\right)$$

and formula (2.1) follows by renaming p_1 to p in the first term and observing that the rest of the right hand side is the sum of all natural numbers from 1 up to $\pi(\sqrt{x}) - 1$. □

Formula (2.1) gives an expression of $\pi_2(x)$ in terms of the prime counting function $\pi(x)$, which can be approximated using the prime number theorem [**21**] as

$$\pi(x) = \alpha_n(x) + \mathcal{O}\left(\frac{x}{(\log x)^{n+1}}\right), \tag{2.2}$$

where $n \in \mathbb{N}$ and

$$\alpha_n(x) = \frac{x}{\log x}\left(\sum_{i=0}^{n-1} \frac{i!}{(\log x)^i}\right). \tag{2.3}$$

Using Landau [**15**, §56], we can rewrite equation (1.6) for any integer $n \in \mathbb{N}$ as

$$\sum_{p \leq \sqrt{x}} \frac{1}{p} = \log(\log x) - \log 2 + M + o\left(\frac{1}{(\log x)^n}\right), \tag{2.4}$$

where we use the little o asymptotic notation [**5**], as opposed to the big $\mathcal{O}$ asymptotic notation used in the prime number theorem (2.2). First we use this result to approximate the sum on the right hand side of equation (2.1).

LEMMA 2.2. *Let $n \in \mathbb{N}$ and $\alpha_n(x)$ be defined by* (2.3). *Then we have*

$$\sum_{p \le \sqrt{x}} \pi\left(\frac{x}{p}\right) = \sum_{p \le \sqrt{x}} \alpha_n\left(\frac{x}{p}\right) + \mathcal{O}\left(\frac{x \log(\log x)}{(\log x)^{n+1}}\right). \tag{2.5}$$

Proof. Using equation (2.2), we have

$$\left|\sum_{p \le \sqrt{x}} \pi\left(\frac{x}{p}\right) - \sum_{p \le \sqrt{x}} \alpha_n\left(\frac{x}{p}\right)\right| \le \sum_{p \le \sqrt{x}} \left|\pi\left(\frac{x}{p}\right) - \alpha_n\left(\frac{x}{p}\right)\right| \le c\,x \sum_{p \le \sqrt{x}} \frac{1}{p\,(\log x - \log p)^{n+1}}\,,$$

where $c > 0$ is a constant. Equation (2.5) then follows by estimating the right hand side by

$$\frac{c\,x\,2^{n+1}}{(\log x)^{n+1}} \sum_{p \le \sqrt{x}} \frac{1}{p} = \mathcal{O}\left(\frac{x \log(\log x)}{(\log x)^{n+1}}\right),$$

where the last equality follows from equation (2.4). □

2.1. *The first two terms of the asymptotic series for $\pi_2(x)$*

Using (2.5) for $n = 1$, we obtain

$$\sum_{p \le \sqrt{x}} \pi\left(\frac{x}{p}\right) = \sum_{p \le \sqrt{x}} \frac{x}{p\,(\log x - \log p)} + \mathcal{O}\left(\frac{x \log(\log x)}{(\log x)^2}\right). \tag{2.6}$$

Using Landau [**15**, §56], we can rewrite equation (1.8) for any integers $i \in \mathbb{N}$ and $n \in \mathbb{N}$ as

$$\sum_{p \le \sqrt{x}} \frac{(\log p)^i}{p} = \frac{(\log x)^i}{i\,2^i} + B_i + o\left(\frac{1}{(\log x)^n}\right). \tag{2.7}$$

We will use this to prove the first theorem of this section.

THEOREM 2.3. *Let M be the Meissel–Mertens constant defined by* (1.6). *Then*

$$\pi_2(x) = \frac{x \log(\log x)}{\log x} + M \frac{x}{\log x} + o\left(\frac{x}{\log x}\right). \tag{2.8}$$

Proof. Using equations (2.1), (2.2) for $n = 1$, and (2.6), we obtain

$$\pi_2(x) = \sum_{p \le \sqrt{x}} \pi\left(\frac{x}{p}\right) + o\left(\frac{x}{\log x}\right) = \frac{x}{\log x} \sum_{p \le \sqrt{x}} \frac{\log x}{p\,(\log x - \log p)} + o\left(\frac{x}{\log x}\right). \tag{2.9}$$

We have the following identity

$$\frac{\log x}{p\,(\log x - \log p)} = \frac{1}{p} + \frac{\log p}{\log x}\left(\frac{\log x}{p\,(\log x - \log p)}\right).$$

Substituting the left hand side into the right hand side, we obtain, for any natural number $n \in \mathbb{N}$, that

$$\frac{\log x}{p\,(\log x - \log p)} = \frac{1}{p} \sum_{i=0}^{n} \left(\frac{\log p}{\log x}\right)^i + \left(\frac{\log p}{\log x}\right)^{n+1} \left(\frac{\log x}{p\,(\log x - \log p)}\right).$$

Summing over all primes $p \le \sqrt{x}$, we get

$$\sum_{p \le \sqrt{x}} \frac{\log x}{p\,(\log x - \log p)} = \sum_{i=0}^{n} \left(\frac{1}{(\log x)^i} \sum_{p \le \sqrt{x}} \frac{(\log p)^i}{p}\right) + \frac{1}{(\log x)^n} \sum_{p \le \sqrt{x}} \frac{(\log p)^{n+1}}{p\,(\log x - \log p)}.$$

Each term on the right hand side can be evaluated using equations (2.4) and (2.7), with $o(1)$ accuracy, as

$$\sum_{p\le\sqrt{x}} \frac{\log x}{p\,(\log x-\log p)} = \log(\log x) - \log 2 + M + \sum_{i=1}^{n}\frac{1}{i\,2^i} + f(n) + o(1), \tag{2.10}$$

where $f(n)$ is a decreasing function of n satisfying $f(n)\to 0$ as $n\to\infty$. This can be deduced by applying equation (2.7) to the error term

$$\frac{1}{(\log x)^n}\sum_{p\le\sqrt{x}}\frac{(\log p)^{n+1}}{p\,(\log x-\log p)} \le \frac{2}{(\log x)^{n+1}}\sum_{p\le\sqrt{x}}\frac{(\log p)^{n+1}}{p} = \frac{1}{(n+1)\,2^n} + o(1).$$

Since we have

$$-\log 2 + \sum_{i=1}^{n}\frac{1}{i\,2^i} + f(n) \to 0, \qquad \text{as} \quad n\to\infty,$$

equation (2.10) implies that

$$\sum_{p\le\sqrt{x}} \frac{\log x}{p\,(\log x-\log p)} = \log(\log x) + M + o(1).$$

Substituting into equation (2.9), we obtain formula (2.8). □

2.2. *Asymptotic series for the counting function of semiprimes*

To derive formulas for all terms in the asymptotic series of π_2, we first define an auxiliary sequence of numbers q_n for $n\in\mathbb{N}$ by

$$q_n = \sum_{i=1}^{n-1}\frac{2^i-1}{i}, \quad \text{for } n\ge 2, \qquad \text{and} \qquad q_1 = 0\,. \tag{2.11}$$

Then q_n is an increasing sequence of rational numbers with the first few terms given as $q_1=0$, $q_2=1$, $q_3=5/2$, $q_4=29/6$, $q_5=103/12$ and $q_6=887/60$, which satisfies the following identity.

Lemma 2.4. *Let $n\in\mathbb{N}$ and let q_n be given by equation* (2.11). *Then we have*

$$\sum_{i=1}^{\infty}\binom{n+i-1}{n-1}\frac{1}{i\,2^i} = q_n + \log 2\,. \tag{2.12}$$

Proof. Considering the binomial series

$$(1-t)^{-n} = 1 + t\sum_{i=0}^{\infty}\binom{n+i}{n-1}t^i, \qquad \text{for } t\in(-1,1),$$

we can rewrite it as

$$\sum_{i=0}^{\infty}\binom{n+i}{n-1}t^i = \frac{(1-t)^{-n}-1}{t} = \sum_{i=1}^{n}(1-t)^{-i}.$$

Integrating, we get

$$\sum_{i=0}^{\infty}\binom{n+i}{n-1}\frac{t^{i+1}}{i+1} = -\log(1-t) + \sum_{i=2}^{n}\frac{(1-t)^{-i+1}-1}{i-1},$$

which holds for $t\in(-1,1)$. Substituting $t=1/2$, we obtain (2.12). □

We will use Lemma 2.4 in the proof of the following theorem, giving the asymptotic series for the semiprime counting function $\pi_2(x)$.

THEOREM 2.5. *The constants C_n appearing in the asymptotic expansion* (1.5) *are given by equation* (1.7) *for $n \in \mathbb{N}$ and as $C_0 = B_0 = M$ for $n = 0$.*

Proof. The case $n = 0$ is studied in Theorem 2.3, which states that $C_0 = M$. To derive equation (1.7), we again use formula (2.1) from Lemma 2.1 and approximate each term using the prime number theorem (2.2). We need to analyze sums of the form

$$S_n(x) = \sum_{p \le \sqrt{x}} \frac{(\log x)^n}{p\,(\log x - \log p)^n} = \sum_{p \le \sqrt{x}} \frac{1}{p}\left(1 - \frac{\log p}{\log x}\right)^{-n}. \tag{2.13}$$

Using the binomial series on the right hand side, we get

$$S_n(x) = \sum_{p \le \sqrt{x}} \frac{1}{p} \sum_{i=0}^{\infty} \binom{n+i-1}{n-1} \left(\frac{\log p}{\log x}\right)^i = \sum_{i=0}^{\infty} \binom{n+i-1}{n-1} \frac{1}{(\log x)^i} \sum_{p \le \sqrt{x}} \frac{(\log p)^i}{p}. \tag{2.14}$$

Substituting x^2 for x, we obtain

$$S_n(x^2) = \sum_{i=0}^{\infty} \binom{n+i-1}{n-1} \frac{1}{2^i (\log x)^i} \sum_{p \le x} \frac{(\log p)^i}{p}. \tag{2.15}$$

To estimate the sums over primes on the right hand side, we apply the result of Rosser and Schoenfeld [**18**, equation (2.26)], which can be formulated as

$$\sum_{p \le x} \frac{1}{p} = \log(\log x) + \mathcal{L}_0(x), \qquad \sum_{p \le x} \frac{(\log p)^i}{p} = \frac{(\log x)^i}{i} + \mathcal{L}_i(x),$$

where the error terms $\mathcal{L}_i(x)$ are defined by

$$\begin{aligned}
\mathcal{L}_0(x) &= -\log(\log 2) + \frac{\mathrm{li}(2)}{2} + \frac{\pi(x) - \mathrm{li}(x)}{x} + \int_2^x \frac{\pi(y) - \mathrm{li}(y)}{y^2}\,\mathrm{d}y, \\
\mathcal{L}_i(x) &= -\frac{(\log 2)^i}{i} + \frac{(\log 2)^i\,\mathrm{li}(2)}{2} + \frac{(\log x)^i}{x}\big(\pi(x) - \mathrm{li}(x)\big) \\
&\quad + \int_2^x \frac{(\log y - i)\,(\log y)^{i-1}}{y^2}\big(\pi(y) - \mathrm{li}(y)\big)\,\mathrm{d}y, \qquad \text{for } i \in \mathbb{N}.
\end{aligned}$$

Using this notation and identity (2.12) in Lemma 2.4, we rewrite equation (2.15) as

$$S_n(x^2) = \log(\log x) + q_n + \log(2) + \sum_{i=0}^{\infty} \binom{n+i-1}{n-1} \frac{1}{2^i} \frac{\mathcal{L}_i(x)}{(\log x)^i}. \tag{2.16}$$

Using the inequality (1.1), there exist constants d_1 and d_2 such that

$$\begin{aligned}
\left|\frac{\mathcal{L}_i(x)}{(\log x)^i}\right| &< \frac{2}{(\log x)^i} + \frac{|\pi(x) - \mathrm{li}(x)|}{x} + \frac{1}{(\log x)^i} \int_2^x \frac{|i - \log y|\,(\log y)^{i-1}}{y^2} \big|\pi(y) - \mathrm{li}(y)\big|\,\mathrm{d}y \\
&\le \frac{2}{(\log x)^i} + d_1 \frac{\exp\left(-d_2\sqrt{\log x}\right)}{(\log x)^{3/4}} + \frac{i\,I(i-1,x) + I(i,x)}{(\log x)^i},
\end{aligned} \tag{2.17}$$

where we define (note that we allow the second argument to be ∞ in this definition):

$$I(i,x) = d_1 \int_2^x (\log y)^{i-3/4} \frac{\exp\left(-d_2\sqrt{\log y}\right)}{y}\,\mathrm{d}y\,.$$

Choose $\ell \in \mathbb{N}$. Our goal is to use (2.17) to estimate the rate of convergence of the sum on the right hand of equation (2.16). To do this we first observe that, for $i \geq \ell$, we have

$$\frac{I(i,x)}{(\log x)^i} \leq \frac{d_1}{(\log x)^\ell} \int_2^x (\log y)^{\ell-3/4} \frac{\exp\left(-d_2\sqrt{\log y}\right)}{y} \, \mathrm{d}y \leq \frac{I(\ell,\infty)}{(\log x)^\ell} \, . \tag{2.18}$$

Using inequalities (2.17) and (2.18) and assuming $\log(x) \geq 1$, we can estimate the remainder of the series on the right hand of equation (2.16) as

$$\begin{aligned} \left| \sum_{i=\ell+1}^{\infty} \binom{n+i-1}{n-1} \frac{1}{2^i} \frac{\mathcal{L}_i(x)}{(\log x)^i} \right| \;\leq\; & \left(\frac{2}{(\log x)^{\ell+1}} + d_1 \frac{\exp\left(-d_2\sqrt{\log x}\right)}{(\log x)^{3/4}} \right) \sum_{i=\ell+1}^{\infty} \binom{n+i-1}{n-1} \frac{1}{2^i} \\ & + \frac{I(\ell,\infty)}{(\log x)^{\ell+1}} \sum_{i=\ell+1}^{\infty} \binom{n+i-1}{n-1} \frac{i}{2^i} \;+\; \frac{I(\ell+1,\infty)}{(\log x)^{\ell+1}} \sum_{i=\ell+1}^{\infty} \binom{n+i-1}{n-1} \frac{1}{2^i} \, . \end{aligned}$$

Since all three sums on the right hand side converge independently of x, we deduce that the remainder is of the order $\mathcal{O}\big((\log x)^{-(\ell+1)}\big)$. Therefore, equation (2.16) becomes

$$S_n(x^2) = \log(\log x) + q_n + \log(2) + \sum_{i=0}^{\ell} \binom{n+i-1}{n-1} \frac{1}{2^i} \frac{\mathcal{L}_i(x)}{(\log x)^i} + \mathcal{O}\left(\frac{1}{(\log x)^{\ell+1}} \right) .$$

This means that an asymptotic expansion of $S_n(x^2)$ in terms of negative powers of $\log x$ is given by the sum of the asymptotic series of terms in equation (2.15). The same is true for $S_n(x)$ in equation (2.14). Thus, using equations (2.4), (2.7), (2.12) and (2.14), we obtain

$$\begin{aligned} S_n(x) &= \sum_{i=0}^{\ell} \binom{n+i-1}{n-1} \frac{1}{(\log x)^i} \sum_{p \leq \sqrt{x}} \frac{(\log p)^i}{p} + \mathcal{O}\left(\frac{1}{(\log x)^{\ell+1}} \right) \\ &= \log(\log x) + M + q_n + \sum_{i=1}^{\ell} \binom{n+i-1}{n-1} \frac{B_i}{(\log x)^i} + \mathcal{O}\left(\frac{1}{(\log x)^{\ell+1}} \right) . \end{aligned} \tag{2.19}$$

Using equations (2.3) and (2.5), we have

$$\sum_{p \leq \sqrt{x}} \pi\left(\frac{x}{p}\right) + o\left(\frac{x}{(\log x)^\ell} \right) = \sum_{n=1}^{\ell} \sum_{p \leq \sqrt{x}} \frac{x\,(n-1)!}{p\,(\log x - \log p)^n} = \sum_{n=1}^{\ell} (n-1)! \, \frac{x\, S_n(x)}{(\log x)^n} \, ,$$

where we used the definition (2.13) of $S_n(x)$ to get the second equality. Using equation (2.19) and notation $B_0 = M$, we obtain

$$\sum_{p \leq \sqrt{x}} \pi\left(\frac{x}{p}\right) = \sum_{n=1}^{\ell} (n-1)! \, \frac{x \log(\log x)}{(\log x)^n} + \sum_{n=1}^{\ell} (n-1)! \left(q_n + \sum_{i=0}^{n-1} \frac{B_i}{i!} \right) \frac{x}{(\log x)^n} + o\left(\frac{x}{(\log x)^\ell} \right) .$$

Thus, using formula (2.1) and the prime number theorem (2.2), we obtain the asymptotic expansion (1.5), where we have

$$C_n = n! \left(q_{n+1} + \sum_{i=0}^{n} \frac{B_i}{i!} \right) - 2^n \sum_{i=1}^{n} (i-1)! \, (n-i)! \, .$$

This can be further simplified by using definition (2.11) of q_n. We get

$$C_n = n! \left(\sum_{i=1}^{n} \frac{2^i - 1}{i} + \sum_{i=0}^{n} \frac{B_i}{i!} - \frac{2^n}{n} \sum_{i=0}^{n-1} \frac{1}{\binom{n-1}{i}} \right) = n! \left(\sum_{i=1}^{n} \frac{2^i - 1}{i} + \sum_{i=0}^{n} \frac{B_i}{i!} - \sum_{i=1}^{n} \frac{2^i}{i} \right) .$$

Subtracting the first and the third sum, we obtain (1.7). □

3. *Computing the constants*

In this section, we use a fast converging series to determine the values of the constants B_n and, as a result, of the constants C_n given by equation (1.7). The first constant, $B_0 = C_0 = M$, is the well-studied Meissel–Mertens constant, so we will focus on constants B_n in the case $n \geq 1$. They have been defined by equations (1.8) or (2.7), which can be rewritten as

$$B_n = \lim_{x\to\infty} \left(\sum_{p\leq\sqrt{x}} \frac{(\log p)^n}{p} - \frac{(\log x)^n}{n\,2^n} \right) = \lim_{x\to\infty} \left(\sum_{p\leq x} \frac{(\log p)^n}{p} - \frac{(\log x)^n}{n} \right). \tag{3.1}$$

To derive a formula for evaluating B_n on a computer, we use the prime zeta function [**12**] defined by

$$P(s) = \sum_p \frac{1}{p^s}\,, \qquad \text{for } s \in (1,\infty). \tag{3.2}$$

Differentiating equation (3.2), we get the formula for the n-th derivative of the prime zeta function as

$$P^{(n)}(s) = (-1)^n \sum_p \frac{(\log p)^n}{p^s} = (-1)^n \sum_{p\leq x} \frac{(\log p)^n}{p^s} + (-1)^n \sum_{p>x} \frac{(\log p)^n}{p^s}. \tag{3.3}$$

The prime zeta function $P(s)$ can also be related to the Riemann zeta function $\zeta(s)$ through the formula [**12**]

$$P(s) = \sum_{i=1}^{\infty} \mu(i)\, \frac{\log\big(\zeta(i\,s)\big)}{i}\,, \qquad \text{for } s \in (1,\infty),$$

where $\mu(n)$ is the Möbius function. Taking the derivative of order n of this expression, we obtain

$$P^{(n)}(s) = \sum_{i=1}^{\infty} \mu(i)\, i^{n-1} \left(\frac{\zeta'}{\zeta}\right)^{(n-1)} (i\,s)\,.$$

Substituting into (3.3), we obtain

$$\sum_{p\leq x} \frac{(\log p)^n}{p^s} = (-1)^n \sum_{i=1}^{\infty} \mu(i)\, i^{n-1} \left(\frac{\zeta'}{\zeta}\right)^{(n-1)} (i\,s) \;-\; \sum_{p>x} \frac{(\log p)^n}{p^s}\,. \tag{3.4}$$

Using integration by parts, we obtain

$$\int_1^{\infty} \frac{(\log u)^{n-1}}{u^s}\,\mathrm{d}u = \frac{(n-1)!}{(s-1)^n} = -\left(\frac{(-1)^n}{s-1}\right)^{(n-1)}.$$

Thus, the second sum on the right hand side of equation (3.4) can be approximated by

$$\sum_{p>x} \frac{(\log p)^n}{p^s} = \int_x^{\infty} \frac{(\log u)^{n-1}}{u^s}\,\mathrm{d}u + o(1) = -\left(\frac{(-1)^n}{s-1}\right)^{(n-1)} - \int_1^{x} \frac{(\log u)^{n-1}}{u^s}\,\mathrm{d}u + o(1)\,.$$

Substituting into equation (3.4), we get

$$\sum_{p\leq x} \frac{(\log p)^n}{p^s} - \int_1^{x} \frac{(\log u)^{n-1}}{u^s}\,\mathrm{d}u = (-1)^n \sum_{i=1}^{\infty} \mu(i)\, i^{n-1} \left(\frac{\zeta'}{\zeta}\right)^{(n-1)} (i\,s) \;+\; \left(\frac{(-1)^n}{s-1}\right)^{(n-1)} + o(1)\,.$$

Taking the limit as $s \to 1$ and substituting into equation (3.1), we obtain

$$B_n = (-1)^n \sum_{i=2}^{\infty} \mu(i)\, i^{n-1} \left(\frac{\zeta'}{\zeta}\right)^{(n-1)} (i) \;+\; (-1)^n \lim_{s\to 1} \left(\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}\right)^{(n-1)}, \tag{3.5}$$

n	B_n	C_n
0	0.26149721284764278375	0.26149721284764278375
1	−1.3325822757332208817	−2.0710850628855780875
2	−2.5551076154464547041	−7.6972777412176108802
3	−10.253827096911327612	−35.345660320564161516
4	−59.332397971808450296	−206.71503925406509339
5	−453.62459086132753356	$-1.5111997871316530251 \times 10^3$
6	$-4.3591249600559955673 \times 10^3$	$-1.3546323682845914021 \times 10^4$
7	$-5.0684840978914262902 \times 10^4$	$-1.4622910675883565523 \times 10^5$
8	$-6.9270677393697978276 \times 10^5$	$-1.8675796280076650637 \times 10^6$
9	$-1.0884508606344556845 \times 10^7$	$-2.7733045258413542557 \times 10^7$
10	$-1.9329009099289751454 \times 10^8$	$-4.7098342357703294361 \times 10^8$

TABLE 1. *Table of constants B_n and C_n, for $n = 0, 1, 2, \ldots, 10$, defined by equations* (2.7) *and* (1.7), *which appear in the asymptotic expansion of $\pi_2(x)$. The values of constants B_n are computed by formula* (3.5) *using the Laurent series* (3.6). *The values of constants C_n are computed by equation* (1.7).

where the first term on the right hand side is a quickly converging series and the limit in the second term can be evaluated using the Laurent expansion of $\zeta(s)$ around $s = 1$. This is given by

$$\zeta(s) = \frac{1}{s-1} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \gamma_n (s-1)^n,$$

where the Stieltjes constants γ_n are computed to 20 significant digits in [**3**]. Then, the Laurent expansion of the logarithmic derivative [**4**] of the Riemann zeta function is

$$\frac{\zeta'(s)}{\zeta(s)} = -\frac{1}{s-1} + \gamma_0 + (-2\gamma_1 - \gamma_0^2)(s-1) + \left(\frac{3}{2}\gamma_2 + 3\gamma_0\gamma_1 + \gamma_0^3\right)(s-1)^2 + \left(-\frac{2}{3}\gamma_3 - 2\gamma_0\gamma_2 - 2\gamma_1^2 - 4\gamma_0^2\gamma_1 - \gamma_0^4\right)(s-1)^3 + \ldots. \tag{3.6}$$

Constants B_n computed by formula (3.5) using the Laurent series (3.6) are presented in Table 1. Once we know constants B_n to the desired accuracy, we can use equation (1.7) to calculate constants C_n. They are also presented in Table 1 to 20 significant digits.

4. *Computational results: behaviour of error terms*

In this section, we illustrate the accuracy of the asymptotic series (1.5) by calculating its error terms at each order. Since (1.5) is a sum of two formal asymptotic series, we have two ways to define its errors. First, we can truncate both sums after the same number, ℓ, of terms to define the relative error

$$\varepsilon_{2\ell}(x) = \frac{1}{\pi_2(x)} \left| \left(\sum_{n=1}^{\ell} (n-1)! \frac{x \log(\log x)}{(\log x)^n} + \sum_{n=1}^{\ell} C_{n-1} \frac{x}{(\log x)^n} \right) - \pi_2(x) \right|, \quad \text{for } \ell \in \mathbb{N}, \tag{4.1}$$

which is plotted in Figure 1(a) for $x = 10^6$, $x = 10^8$ and $x = 10^{10}$. Combining both sums in the asymptotic expansion (1.5) into one, we can write it as

$$\pi_2(x) \sim \frac{x \log(\log x)}{\log x} + C_0 \frac{x}{\log x} + \frac{x \log(\log x)}{(\log x)^2} + C_1 \frac{x}{(\log x)^2} + \frac{2\, x \log(\log x)}{(\log x)^3} + \ldots,$$

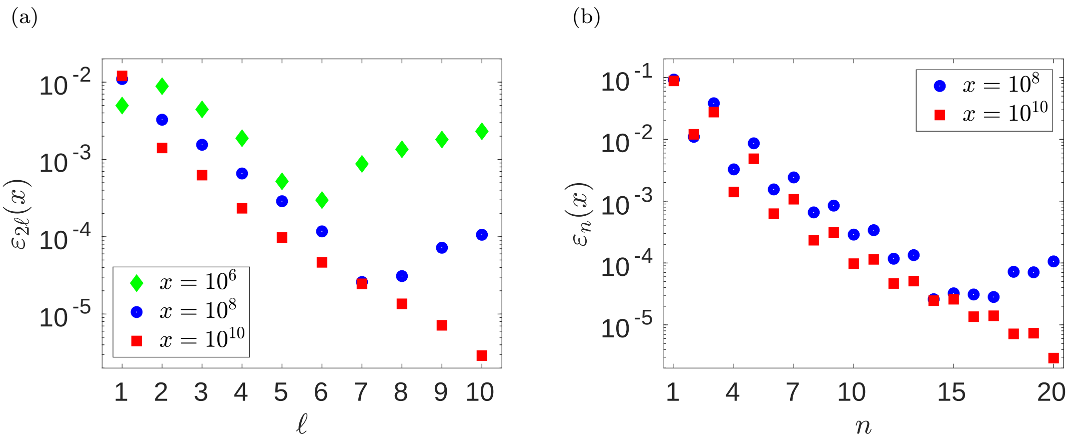

FIGURE 1. *(a) Relative errors calculated by* (4.1) *for* $x = 10^6$, $x = 10^8$ *and* $x = 10^{10}$. *(b) Relative errors calculated by* (4.5) *for* $x = 10^8$ *and* $x = 10^{10}$. *The relative error of Landau's approximation* (4.2) *is plotted as* $\varepsilon_1(x)$. *The relative error of the approximation given in Theorem* 2.3 *is plotted as* $\varepsilon_2(x)$.

so we can define the n-th approximation, $a_n(x)$, by

$$a_1(x) = \frac{x \log(\log x)}{\log x}, \tag{4.2}$$

$$a_{2\ell}(x) = \sum_{i=1}^{\ell} (i-1)! \, \frac{x \log(\log x)}{(\log x)^i} + \sum_{i=1}^{\ell} C_{i-1} \frac{x}{(\log x)^i}, \qquad \text{for } \ell = 1,\, 2,\, 3,\, \dots, \tag{4.3}$$

$$a_{2\ell+1}(x) = \sum_{i=1}^{\ell+1} (i-1)! \, \frac{x \log(\log x)}{(\log x)^i} + \sum_{i=1}^{\ell} C_{i-1} \frac{x}{(\log x)^i}, \qquad \text{for } \ell = 1,\, 2,\, 3,\, \dots, \tag{4.4}$$

where approximation $a_1(x)$ is used in equation (1.3) and $\alpha_2(x)$ is the approximation given in equation (2.8). With definitions (4.2)–(4.4), we can define the relative error by

$$\varepsilon_n(x) = \frac{|a_n(x) - \pi_2(x)|}{\pi_2(x)}. \tag{4.5}$$

This definition is consistent with (4.1) for even values of n. The results computed by (4.5) are plotted in Figure 1(b) for $x = 10^8$ and $x = 10^{10}$.

In Figure 1, we study the behaviour of the relative error (4.5) for a fixed value of x. We observe that the relative error $\varepsilon_n(x)$ is initially a decreasing sequence (for smaller values of n). It reaches its minimum and then it starts increasing again. Denoting the value of n where the relative error reaches its minimum as $n_{\min}(x)$, we observe that $n_{\min}(x)$ is an increasing function of x, at least for the three values of x considered in Figure 1(a).

Since x is fixed in Figure 1, the relative error (4.5) is a constant multiple of the absolute error $|a_n(x) - \pi_2(x)|$. In particular, the same conclusion about $n_{\min}(x)$ could be reached when considering the absolute errors instead of relative errors. Since we plot results for two or three different (fixed) values of x in Figure 1, we can also observe that the relative error is (for many of the values of n) a decreasing function of x, although this does not look to be true for $n = 2$ in Figure 1.

To confirm this observation, we plot the behaviour of the relative error $\varepsilon_n(x)$ as a function of x in Figure 2. We use the first eight values of n and confirm that the relative error $\varepsilon_n(x)$ resembles a decreasing function of x for $n = 1,\, 3,\, 4,\, \dots,\, 8$, while $\varepsilon_2(x)$ reaches its minimum

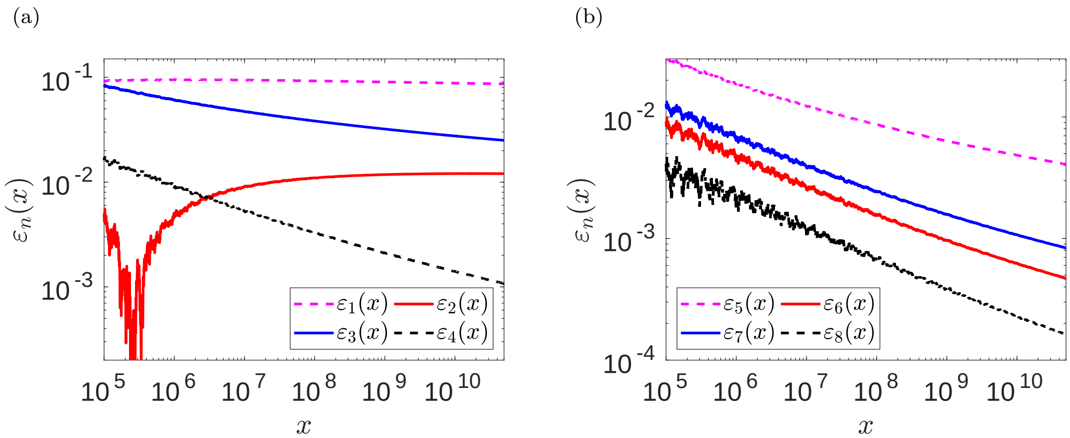

FIGURE 2. *(a) Relative errors $\varepsilon_n(x)$, for $n = 1$, 2, 3 and 4, calculated by* (4.5) *as a function of x. The relative error of Landau's approximation* (4.2) *is plotted as $\varepsilon_1(x)$. The relative error of the approximation given in Theorem* 2.3 *is plotted as $\varepsilon_2(x)$.*
(b) Relative errors $\varepsilon_n(x)$, for $n = 5$, 6, 7 and 8 as a function of x.

between 2×10^5 and 3×10^5. For smaller values of x, the approximation $a_2(x)$ overestimates $\pi_2(x)$, while it is an underestimate for larger values of x. The graphs of $a_2(x)$ and $\pi_2(x)$ cross for the values of x between 2×10^5 and 3×10^5, so in this interval we can get very close to the correct answer, which results in the minimum of $\varepsilon_2(x)$ in Figure 2(a), because our error definition (4.5) includes the absolute value. In Figure 2, we also observe that the other errors are not strictly decreasing, but they fluctuate with a decreasing trend, see, for example, the plot of $\varepsilon_8(x)$ in Figure 2(b).

5. *Discussion*

In this paper, we have studied the behaviour of the semiprime counting function $\pi_2(x)$, which is a special case ($k = 2$) of the k-almost prime counting function $\pi_k(x)$. To generalize the presented results to the case $k \geq 3$, we need to first generalize the counting Lemma 2.1. Using the inclusion-exclusion principle, it is possible to deduce the following counting formula

$$\pi_k(x) = \sum_{i=1}^{k} (-1)^{i-1} \sum_{p_1 < p_2 < \ldots < p_i \leq \sqrt[k]{x}} \pi_{k-i}\left(\frac{x}{p_1 p_2 \ldots p_i}\right), \tag{5.1}$$

where we define function $\pi_0(x)$ to be identically equal to 1, i.e. $\pi_0(x) = 1$, and the sum over $p_1 < p_2 < \ldots < p_i \leq \sqrt[k]{x}$ means that we are summing i-times over all primes satisfying the given condition. Substituting $k = 2$ into equation (5.1), we obtain

$$\pi_2(x) = \sum_{p_1 \leq \sqrt{x}} \pi_1\left(\frac{x}{p_1}\right) - \sum_{p_1 < p_2 \leq \sqrt{x}} \pi_0\left(\frac{x}{p_1 p_2}\right).$$

Using $\pi_0(x) = 1$, we deduce equation (2.1). Thus, equation (5.1) provides a generalization of equation (2.1), which expresses the k-almost prime counting function $\pi_k(x)$ in terms of the counting functions $\pi_1(x)$, $\pi_2(x)$, $\ldots$, $\pi_{k-1}(x)$. It can be inductively used to derive forms of coefficients of polynomials $P_{n,k}$ in the asymptotic series (1.4). In addition to constants B_n and C_n, certain new constants will appear in such calculations, including the (converging) sums of the form $\sum_p (\log p)^i p^{-\ell}$ with $\ell \geq 2$ and $i \in \mathbb{N}$. For a detailed discussion of the asymptotic behaviour of these sums for $\ell = 1$, see Axler [**2**]. Substituting $n = \pi(x)$ in [**2**, Theorem 5] gives a different expansion for the sums in (1.8), which may be further examined using the prime number theorem.

There are, also, other possible approximations for $\pi_k(x)$. For example, Erdős and Sárközy [**10**] prove that

$$\pi_k(x) < \begin{cases} c(\delta)\, \dfrac{x}{\log x}\, \dfrac{(\log(\log x))^{k-1}}{(k-1)!}\,, & \text{for } 1 \le k \le (2-\delta)\log(\log x); \\ c\, k^4\, 2^{-k}\, x\, \log x\,, & \text{for } k \ge 1; \end{cases}$$

for some constants $c(\delta)$ and c. Other approximations, relating the function π_k to some other products over primes are possible to obtain, as explained in [**21**].

Functions $\pi_k(x)$ and $\Omega(x)$, used in expansion (1.9), count the prime divisors with their multiplicity. Another possible generalization is to investigate the related functions $N_k(x)$ and $\omega(x)$, counting prime divisors without multiplicity. That is, functions $N_k(x)$ and $\omega(x)$ are defined to be the number of natural numbers $n \le x$ which have exactly k distinct prime divisors and the number of distinct prime divisors of x, respectively. Finch [**11**, page 26] shows that

$$\frac{1}{x} \sum_{m \le x} \omega(m) \;\sim\; \log(\log x) + 0.2614972128\cdots + \sum_{n=1}^{\infty} \left(-1 + \sum_{i=0}^{n-1} \frac{\gamma_i}{i!} \right) \frac{(n-1)!}{\log^n x}\,,$$

which has the higher order terms in the same form as in the expansion (1.9). Using the prime number theorem, we also observe that $N_1(x) = \pi\,(x) + \pi\,(\sqrt{x}) + \pi\,(\sqrt[3]{x}) + \ldots \sim \mathrm{li}(x)$ admits an identical asymptotic expansion as $\pi(x)$. Delange [**6**, Theorem 1] and Tenenbaum [**21**] obtained the asymptotic expansion of $N_k(x)$ in the form

$$N_k(x) \;\sim\; \frac{x}{\log x} \sum_{n=0}^{\infty} \frac{Q_{n,k}(\log(\log x))}{(\log x)^n}\,,$$

where $Q_{n,k}$ are polynomials of degree $k-1$. Here, the expansion is similar to the expansion (1.4) for $\pi_k(x)$, but the polynomials $P_{n,k}$ and $Q_{n,k}$ are different. Results about the leading terms of polynomials $Q_{n,k}$ and about $Q_{0,k}$ have also been obtained, as in the case of π_k. Several different approximations for $N_k(x)$ are also possible to derive, as shown in Tenenbaum [**21**], who points out that the function N_k is easier to analyse than π_k, for larger values of k, relative to $\log(\log x)$. For example, the following holds uniformly for $x \ge 3$ and $(2+\delta)\log(\log x) \le k \le A\log(\log x)$:

$$N_k(x) = C\, \frac{x \log x}{2^k} \left(1 + \mathcal{O}\Big((\log x)^{-\delta^2/5} \Big) \right)\,,$$

where $A > 0$, $0 < \delta < 1$ and $C \approx 0.378694$. Similar results, but for larger values of k, can be obtained for functions π_k as well [**16**].

Acknowledgements. Authors would like to thank Julia Stadlmann for helpful comments during the preparation of this manuscript.

References

1. C. Axler. New bounds for the prime counting function. *Integers*, 16:A22, 2006.
2. C. Axler. On a family of functions defined over sums of primes. *Journal of Integer Sequences*, 22(5):Article 19.5.7, 2019.
3. J. Bohman and C. Fröberg. The Stieltjes function – definition and properties. *Mathematics of Computation*, 51(163):281–289, 1988.
4. K. Broughan. Extension of the Riemann ζ-functions logarithmic derivative positivity region to near the critical strip. *Canadian Mathematical Bulletin*, 52(2):186–194, 2009.
5. N. de Bruijn. *Asymptotic Methods in Analysis*. North-Holland Publishing Co., Amsterdam, 1958.
6. H. Delange. Sur des formules de Atle Selberg. *Acta Arithmetica*, 19:105–146, 1971.

7. P. Diaconis. Asymptotic expansion for the mean and variance of the number of prime factors of a number n. Technical Report 96, Department of Statistics, Stanford University, Stanford, California, December 1976.
8. P. Dusart. Explicit estimates of some functions over primes. *Ramanujan Journal*, 45:227–251, 2018.
9. P. Erdős and M. Kac. The Gaussian law of errors in the theory of additive number theoretic functions. *American Journal of Mathematics*, 62(1):738–742, 1940.
10. P. Erdős and A. Sárkőzy. On the number of prime factors of integers. *Acta Scientiarum Mathematicarum*, 42:237–246, 1980.
11. S. Finch. *Mathematical Constants II.* Cambridge University Press, Cambridge, 2018.
12. C. Fröberg. On the prime zeta function. *BIT*, 8:187–202, 1968.
13. A. Harper. Two new proofs of the Erdős-Kac theorem, with bound on the rate of convergence, by Stein's method for distributional approximations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 147(1):95114, 2009.
14. S. Ishmukhametov and F. Sharifullina. On distribution of semiprime numbers. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, 8:53–59, 2014.
15. E. Landau. *Handbuch der Lehre von der Verteilung der Primzahlen.* Druck und Verlag von B.G.Teubner, Lepzig und Berlin, 1909.
16. J. Nicolas. Sur la distribution des nombres entiers ayant une quantité fixée de facteurs premiers. *Acta Arithmetica*, 44(3):191–200, 1984.
17. A. Rényi and P. Turán. On a theorem of Erdős-Kac. *Acta Arithmetica*, 4(1):71–84, 1958.
18. J. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, 6(1):64–94, 1962.
19. J. Rosser and L. Schoenfeld. Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. *Mathematics of Computation*, 29(129):243–269, 1975.
20. L. Schoenfeld. Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II. *Mathematics of Computation*, 30(134):337–360, 1976.
21. G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*, volume 163 of *Graduate Studies in Mathematics*. American Mathematical Society, 2015.
22. T. Trudgian. Updating the error term in the prime number theorem. *Ramanujan Journal*, 39:225–234, 2016.

Dragoș Crișan
Merton College
Merton Street
Oxford OX1 4JD
United Kingdom

dragos.crisan@merton.ox.ac.uk

Radek Erban
Mathematical Institute
University of Oxford
Woodstock Road
Oxford OX2 6GG
United Kingdom

erban@maths.ox.ac.uk