

*The local distribution of the number
of small prime factors: variation of the
classical theme*

Krishnaswami Alladi & Todd Molnar

The Ramanujan Journal

An International Journal Devoted to
the Areas of Mathematics Influenced by
Ramanujan

ISSN 1382-4090

Volume 51

Number 1

Ramanujan J (2020) 51:117-151

DOI 10.1007/s11139-019-00151-2

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



The local distribution of the number of small prime factors: variation of the classical theme

Krishnaswami Alladi¹ · Todd Molnar¹Received: 23 March 2018 / Accepted: 18 February 2019 / Published online: 31 July 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

We obtain uniform estimates for $N_k(x, y)$, the number of positive integers n up to x for which $\omega_y(n) = k$, where $\omega_y(n)$ is the number of distinct prime factors of n which are $< y$. The motivation for this problem is an observation due to the first author in 2015 that for certain ranges of y , the asymptotic behavior of $N_k(x, y)$ is different from the classical situation concerning $N_k(x, x)$ studied by Sathe and Selberg. We demonstrate this variation of the classical theme; to estimate $N_k(x, y)$ we study the sum $S_z(x, y) = \sum_{n \leq x} z^{\omega_y(n)}$ for $\operatorname{Re}(z) > 0$ by the Buchstab–de Bruijn method. We also utilize a certain recent result of Tenenbaum to complete our asymptotic analysis.

Keywords Number of small prime factors · Local distribution · Landau’s theorem · Selberg’s method · Buchstab iteration · de Bruijn’s method · Difference-differential equations

Mathematics Subject Classification 11M06 · 11M41 · 11N25 · 11N37 · 11N60

1 Introduction

The function

$$\omega_y(n) = \sum_{\substack{p|n \\ p < y}} 1,$$

where p is a prime number, figures prominently in the proofs of the Erdős–Kac theorem which concerns the global distribution of the number of prime factors. Here we focus on the “local distribution” of $\omega_y(n)$, that is, we study the function

✉ Krishnaswami Alladi
alladik@ufl.edu

Todd Molnar
twmolnar@gmail.com

¹ Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

$$N_k(x, y) = \sum_{\substack{n \leq x \\ \omega_y(n)=k}} 1$$

with emphasis on results which are uniform. Throughout we will assume that $x, y \geq 3$ to avoid negative values for certain key functions.

The first significant study of $\omega(n)$ is due to Hardy and Ramanujan [9] in 1917. They proved that $\omega(n)$ is almost always asymptotically $\log \log n$ in size by obtaining a uniform bound for the function $N_k(x, x) = N_k(x)$; they showed that there exist absolute constants $A, B > 0$ such that

$$N_k(x) = \sum_{\substack{n \leq x \\ \omega(n)=k}} 1 < \frac{Bx(\log \log x + A)^{k-1}}{(k - 1)! \log x}$$

for all $x \geq 3$ and all integers $k \geq 1$.

When $\alpha = \log(x)/\log(y) > 1$, and k is small, the behavior of $N_k(x, y)$ is different from the classical case of $N_k(x)$ (see Theorems 2, 3, 11, and 12), a phenomenon first observed by Alladi in 2015. We will supply the analysis necessary to explain this phenomenon. However, as k approaches $\log \log y$, or as y approaches x , the behavior is similar to the classical situation. We investigate how and when such a transition takes place.

In preparation for this investigation, we shall study the behavior of the sum

$$S_z(x, y) = \sum_{n \leq x} z^{\omega_y(n)}, \tag{1}$$

where $z \in \mathbb{C}$. Note that in the special case $z = 0, S_0(x, y) = \Phi(x, y)$, the well-known function counting the number of uncanceled elements in the Sieve of Eratosthenes. Obtaining asymptotic estimates for $S_z(x, y)$ which are both sharp enough for application and uniform in y will require a variety of tools. Besides the elementary techniques of Sect. 3, we shall also employ analytic methods in Sects. 2, 4, and 5 such as de Bruijn's method of utilizing Buchstab iterations and difference-differential equations [5]. It will be important to note that the ranges for the estimates in these sections overlap, thereby permitting a result valid in a larger range. More importantly, by comparing the estimates in these two ranges, we obtain an asymptotic estimate for a certain function satisfying a difference-differential equation without going through a saddle point analysis; this technique was used by Alladi in [3].

Once we have asymptotic estimates for $S_z(x, y)$, we may then study $N_k(x, y)$ by recognizing that

$$N_k(x, y) = \frac{1}{2\pi i} \int_{\gamma} \frac{S_z(x, y)}{z^{k+1}} dz \tag{2}$$

for a suitable contour γ . The idea of using such a contour integral is due to Selberg [12] who studied the case $y = x$. If we let

$$g(s, z) = \prod_p \left(1 + \frac{z}{p^s - 1}\right) \left(1 - \frac{1}{p^s}\right)^z$$

then Selberg demonstrated that with $r = Re(z)$

$$S_z(x, x) = S_z(x) = \frac{g(1, z)}{\Gamma(z)} \frac{x}{\log^{1-z} x} + O\left(\frac{x}{\log^{2-r} x}\right)$$

uniformly for $|z| \leq R$, where $\Gamma(z)$ is the Gamma function. Then using the integral in (2) with γ a circle centered at the origin with radius

$$\rho = \frac{k - 1}{\log \log x},$$

Selberg demonstrated that

$$N_k(x) = \frac{x}{\log x} \frac{g\left(1, \frac{k-1}{\log \log x}\right)}{\Gamma\left(1 + \frac{k-1}{\log \log x}\right)} \frac{(\log \log x)^{k-1}}{(k - 1)!} \left(1 + O\left(\frac{k}{(\log \log x)^2}\right)\right) \quad (3)$$

uniformly for $k \leq R \log \log x$, where R is any fixed positive number. This result improves upon that of Landau (see (7) below) and Sathe in [11]. Our methods leading to Theorems 10 and 12 show that for certain ranges of k and y

$$N_k(x, y) \asymp N_{k+1}(x);$$

in such a situation, the function $N_k(x, y)$ is of the size of the $(k + 1)$ -st Landau function. The above equation is the precise statement of the phenomenon observed by Alladi in 2015, and our results pertaining to this were established in 2016.

In a fundamental paper Halász [7] studied the local distribution of the general additive function

$$\omega(n; Q) := \sum_{\substack{p|n \\ p \in Q}} 1,$$

where Q is an arbitrary set of prime numbers. He obtained the asymptotic estimate

$$N_k(x; Q) := \sum_{\substack{n \leq x \\ \omega(n; Q) = k}} 1 \sim \frac{x e^{-Q(x)} Q(x)^k}{k!}$$

when $\omega(n; Q) \sim Q(x)$, where

$$Q(x) = \sum_{\substack{p \leq x \\ p \in E}} \frac{1}{p}.$$

(see Elliott [6, vol. 2, pp. 296–301] for a proof). In a recent paper Tenenbaum [15] has strengthened Halász’s result by extending the range of Halász’ asymptotic formula to

$$1/\kappa \leq k/Q(x) \leq \kappa$$

for any large $\kappa > 0$; his results are uniform in that range. In the course of explaining the phenomenon observed by Alladi, Tenenbaum [14] communicated to us in 2016 that by choosing $Q = \{p|p < y\}$, one may use the Selberg–Delange method and other techniques to obtain sharper error terms than ours in some instances (see [17]). The emphasis of this paper is to obtain sufficiently sharp and uniform asymptotic estimates that will demonstrate and explain the phenomenon observed by Alladi, which seems to have escaped attention. Our methods are different from those of Halász and Tenenbaum and are of intrinsic interest.

Some Notation: We use the standard notation in number theory $s = \sigma + it$ for a complex number s when we discuss analytic functions of s in right half planes. But for the complex number z occurring in the definition of $S_z(x, y)$, we will put $r = Re(z)$.

Throughout we assume that $x, y \geq 3$ so that $\log \log x$ and $\log \log y$ are positive. We use the logarithmic integral defined as

$$li(y) = \int_2^y \frac{dt}{\log t}.$$

We let $\pi(y)$ denote the number of primes up to y . With $li(y)$ as above, we denote by $E(y)$ any decreasing function of y that bounds from above the relative error in the Prime Number Theorem

$$\frac{|\pi(y) - li(y)|}{li(y)} < E(y),$$

with $E(y) \log^R y \rightarrow 0$ as $y \rightarrow \infty$, for every $R > 0$.

Also throughout we put $\beta = \frac{x}{y}$ and $\alpha = \frac{\log x}{\log y}$. Further notation will be introduced when needed.

2 Asymptotic estimate of $S_z(x, y)$ for small y

Following the method of de Bruijn in [5] we may obtain an estimate for $S_z(x, y)$ for “small y ” using contour integration. The contour integral representation is valid for any complex $z \in \mathbb{C}$ and will yield an asymptotic estimate of $S_z(x, y)$ provided $z \neq -p + 1$, with p a prime number. First note that as $z^{\omega_y(n)}$ is multiplicative, we have

$$F(s) := \sum_{n=1}^{\infty} \frac{z^{\omega_y(n)}}{n^s} = \zeta(s) \prod_{p < y} \left(1 + \frac{z-1}{p^s}\right) = \zeta(s)g(s, y, z),$$

where $\zeta(s)$ is the Riemann zeta function and

$$g(s, y, z) = \prod_{p < y} \left(1 + \frac{z - 1}{p^s} \right),$$

so that the function $g(s, y, z)$ is now analytic in both s and z .

Lemma 1 *If $\text{Re}(s) = \sigma \geq 1 - \frac{1}{\log y}$ and $|z| \leq R$ then there exists a constant $C > 0$ such that*

$$g(s, y, z) \ll_R \log^{C|z-1|} y.$$

The proof follows using the methods in [5].

With the estimate of Lemma 1 we may now prove

Theorem 1 *Let $R > 0$ be fixed, $|z| \leq R$, and $x \geq y \geq 3$. Then*

$$S_z(x, y) = x \prod_{p < y} \left(1 + \frac{z - 1}{p} \right) + O(xe^{-\alpha} \log^D x) + O_R \left(\frac{x}{\log^{R+2} x} \right)$$

with some constant $D > 0$, and where $\alpha = \frac{\log x}{\log y}$.

Proof From the Perron integral formula (see [13, p. 132]) we have the following representation:

$$\begin{aligned} S_z(x, y) &= \sum_{n \leq x} z^{\omega_y(n)} \\ &= \frac{1}{2\pi i} \int_{a-iT}^{a+iT} \zeta(s) g(s, y, z) \frac{x^s}{s} ds \\ &\quad + O \left(\sum_{n=1}^{\infty} \left(\frac{x}{n} \right)^a (1 + [R])^{\omega(n)} \min \left(1, \frac{1}{T |\log(x/n)|} \right) \right), \end{aligned} \tag{4}$$

where $a > 1$. In (4) we have used the simple upper bound $(1 + [R])^{\omega(n)}$ for the coefficients of the Dirichlet series $F(s)$ when $|s| \leq R$.

Following Selberg [12], the sum in (4) is estimated by dividing the range of summation into three intervals: $n < x/2$, $n > 3x/2$, and $x/2 \leq n \leq 3x/2$. Some care needs to be paid in dealing with n very close to x , and this is why we have replaced R by $1 + [R]$ in the error term above. This yields

$$\sum_{n=1}^{\infty} \left(\frac{x}{n} \right)^a (1 + [R])^{\omega(n)} \min \left(1, \frac{1}{T |\log(x/n)|} \right) \ll_R \frac{x}{\log^{R+2}(x)} + \frac{x \log^{2R}(x)}{T}.$$

By choosing $T := e^{\sqrt{\log(x)}}$ in the above expression, we obtain

$$S_z(x, y) - \frac{1}{2\pi i} \int_{a-iT}^{a+iT} \zeta(s)g(s, y, z) \frac{x^s}{s} ds \ll_R \frac{x}{\log^{R+2}(x)}. \tag{5}$$

To evaluate the integral in Eq. (5), we consider the rectangular contour Γ with vertices $a + iT, b + iT, b - iT$, and $a - iT$ with

$$\max\left(1 - \frac{1}{\log T}, 1 - \frac{1}{\log y}\right) = b < 1 < a = 1 + \frac{1}{\log x}$$

and $T = e^{\sqrt{\log x}}$ defined as above. Using Cauchy's theorem and taking into account the simple pole of $\zeta(s)$ with residue 1 at $s = 1$, we get the term

$$x \prod_{p < y} \left(1 + \frac{z-1}{p}\right)$$

in the Theorem. Upper bounds for the contribution from the vertical line segments of Γ corresponding to $\sigma = b$, and horizontal segments of Γ , yield the O terms in Theorem 1.

Theorem 1 can of course be used to obtain an effective asymptotic estimate for $S_z(x, y)$ for certain ranges of y . In fact, if α is large enough to suppress the second term in Theorem 1 then, provided $y \rightarrow \infty$ and $z \neq 1$ or $1 - p$, the main term will simply be

$$x \prod_{p < y} \left(1 + \frac{z-1}{p}\right) \asymp_z \frac{x}{\log^{1-z} y}.$$

This can be improved with the following □

Corollary 1 *Let $z \in \mathbb{C}$, $|z| \leq R$, and let D denote the constant in Theorem 1. If $\alpha \geq (R + D + 1 + \epsilon) \log \log x$, $\epsilon > 0$ an arbitrary small fixed constant, then*

$$S_z(x, y) = x \prod_{p < y} \left(1 + \frac{z-1}{p}\right) + O\left(\frac{x}{\log^{R+1+\epsilon} y}\right);$$

in particular, there exists a positive constant K such that if $\alpha \geq K \log \log x$ and $z \neq 1$, or $1 - p$ then

$$S_z(x, y) \sim x \prod_{p < y} \left(1 + \frac{z-1}{p}\right).$$

Following de Bruijn in [5] we may introduce an additional parameter $\lambda > 0$ by replacing $1 - \frac{1}{\log y}$ with $1 - \frac{\lambda}{\log y}$, in defining b . Then carrying out the analysis of Theorem 1, we may optimize this parameter to achieve an error term which is

$\ll x e^{-\alpha \log \alpha} \log^D(x)$. We do not require the full strength of this result and so we settled for the simpler bound in Theorem 1. Theorem 1 and Corollary 1 can also be obtained using the more elementary technique of convolution sums, as in Hall and Tenenbaum [8]. Note that if α satisfies the range of Corollary 1, then

$$y = x^{1/\alpha} \leq e^{\frac{\log x}{K \log \log x}}. \tag{6}$$

This range of y will be crucial for subsequent results.

3 Estimate of $N_k(x, y)$ for very large y

In this section, we will derive estimates for $S_z(x, y)$ when y is *very large*, by which we mean that if $\beta = x/y$, then β is fixed. It is a classical result due to Landau [10] (see also Tenenbaum [13, p. 200]) that if k is fixed and $\beta = 1$ then

$$N_k(x) = \frac{x(\log \log x)^{k-1}}{(k-1)! \log x} + O\left(\frac{x(\log \log x)^{k-2}}{(k-2)! \log x}\right). \tag{7}$$

This is implied by (3). We shall see in Theorem 2 that the asymptotic behavior of $N_k(x, y)$ is quite different from (7) for certain ranges of y .

Lemma 2 *For any fixed integer $k > 1$ and $\ell \geq 1$ we have*

$$\sum_{p_1^{e_1} p_2^{e_2} \dots p_k^{e_k} < x} \frac{\log^\ell(p_1^{e_1} \dots p_k^{e_k})}{p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}} = \frac{\log^\ell x (\log \log x)^{k-1}}{(k-1)! \ell} + O\left(\frac{\log^\ell x (\log \log x)^{k-2}}{(k-2)! \ell}\right).$$

Proof We may write

$$\sum_{p_1^{e_1} p_2^{e_2} \dots p_k^{e_k} < x} \frac{\log^\ell(p_1^{e_1} \dots p_k^{e_k})}{p_1^{e_1} p_2^{e_2} \dots p_k^{e_k}} = \int_2^x \frac{\log^\ell(t)}{t} dN_k(t),$$

Lemma 2 follows by using applying integration by parts and using (7). □

Remarks (i) Even though we have stated Lemma 2 for $\ell \geq 1$ and $k \geq 2$ we will use it in the sequel only for $\ell = 1$.

(ii) When $k = \ell = 1$ there is the well-known sharper result

$$\sum_{p^e \leq x} \frac{\log p}{p^e} = \log x + c + O\left(\frac{1}{\log(x)}\right).$$

(iii) The estimate given in Lemma 2 still applies if the summation is over square-free integers, that is, if $\ell \geq 1$ and $k > 1$, then

$$\sum_{p_1 \dots p_k < x} \frac{\log^\ell(p_1 \dots p_k)}{p_1 \dots p_k} = \frac{\log^\ell x (\log \log x)^{k-1}}{(k-1)! \ell} + O\left(\frac{\log^\ell x (\log \log x)^{k-2}}{(k-2)! \ell}\right). \tag{8}$$

To establish (8) we follow the proof given above for Lemma 2 except instead of the integrator $N_k(x)$ we use

$$N_k^*(x) = \sum_{\substack{n \leq x \\ \Omega(n) = \omega(n) = k}} 1,$$

which is the number of square-free integers with precisely k prime factors. The estimates of Lemma 2 and (8) agree because for fixed k

$$N_k(x) = N_k^*(x) + O\left(\frac{x (\log \log x)^{k-2}}{\log x (k-2)!}\right).$$

(iv) When $\ell = 0$, the main term in the asymptotic estimate for the sum in (8) is $(\log \log x)^k / k!$, and for this case, Tenenbaum [16] has recently obtained very precise evaluations.

Let $\beta := x/y$. We assume $x > \beta^2$, so that we can have $\sqrt{x} < y \leq x$. We begin by considering $N_k(x, y)$ when $k = 0$, that is

$$N_0(x, y) = \sum_{\substack{n \leq x \\ \omega_y(n) = 0}} 1 = \sum_{\substack{n \leq x \\ p^-(n) \geq y > \sqrt{x}}} 1 = \Phi(x, y). \tag{9}$$

Since $y > \sqrt{x}$ then the sum in (9) will be counting integers $n \leq x$ with all prime factors $> \sqrt{x}$; however, the only integers with this property are 1 and the prime numbers $y \leq p \leq x$. Hence, we have

Lemma 3 *If $x \geq y > \sqrt{x}$ then*

$$N_0(x, y) = 1 + \pi(x) - \pi(y - 1).$$

From the prime number theorem we may immediately obtain an asymptotic estimate in the form

$$\begin{aligned} N_0(x, y) &= \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right) - \frac{y}{\log y} + O\left(\frac{y}{\log^2 y}\right) \\ &= \frac{x}{\log x} - \frac{y}{\log y} + O\left(\frac{x}{\log^2 x}\right), \end{aligned} \tag{10}$$

with obvious improvements being possible. Comparing this with $N_0(x, x) = 1$ (the corresponding Landau term) we can immediately see that $N_0(x, y)$ is much larger when $\beta > 1$.

Next, with $y > \sqrt{x}$, consider $N_k(x, y)$ when $k = 1$. The numbers which contribute to this sum will be of the form $n = mp^e \leq x$ with $p < y$, $p^-(m) \geq y > \sqrt{x}$, and $m \leq x/p^e$, hence,

$$N_1(x, y) = \sum_{\substack{n \leq x \\ \omega_y(n)=1}} 1 = \sum_{\substack{mp^e \leq x \\ p < y \\ p^-(m) \geq y > \sqrt{x}}} 1 = \sum_{\substack{p^e \leq x \\ p < y}} 1 + \sum_{\substack{m \leq x/p^e \\ p < y \\ p^-(m) \geq y}} 1,$$

where the first term corresponds to $m = 1$ and the second to $m > 1$; however, since $p^-(m) \geq y > \sqrt{x}$ this forces $m = q$ to be prime. Consequently

$$N_1(x, y) = \pi(y - 1) + O(\sqrt{y}) + \sum_{\substack{y \leq q \leq x/p^e \\ p < y}} 1.$$

By evaluating the sum on the right in the above expression, we get

Lemma 4 *If $\sqrt{x} < y \leq x$ with $\beta = x/y$ fixed, then*

$$N_1(x, y) = \frac{x}{\log y} \sum_{p^e < \beta} \left(\frac{1}{p^e} - \frac{1}{\beta} \right) + \frac{x}{\beta \log y} + O_\beta \left(\frac{x}{\log^2 x} \right).$$

Remark In Lemma 4, the expression $\frac{x}{\beta \log y}$ corresponds to main term in Landau's estimate (7) for $k = 1$. Note that this term decreases as β increases.

If as $x \rightarrow \infty$ we now allow β to become large as well, the estimate in Lemma 4 implies that $N_1(x, y) \sim \frac{x \log \log \beta}{\log x}$, as this term dominates the Landau term of $\frac{x}{\beta \log y}$.

We can apply the above techniques (and Lemma 2) to analyze the sum $N_k(x, y)$ for general fixed k and $\beta > 1$.

Theorem 2 *For arbitrary fixed $\beta := x/y > 1$, $\sqrt{x} < y < x$ and $k \in \mathbb{Z}$, $k > 1$ fixed, we have*

$$N_k(x, y) = \frac{x}{\log y} \sum_{p_1^{e_1} \dots p_k^{e_k} \leq \beta} \left(\frac{1}{p_1^{e_1} \dots p_k^{e_k}} - \frac{1}{\beta} \right) + \frac{x(\log \log y)^{k-1}}{\log y(k-1)!} + O \left(\frac{x}{\log y} \frac{(\log \log x)^{k-2}}{(k-2)!} \right).$$

Theorem 2 is proved by noting that

$$N_k(x, y) = \sum_{\substack{p_1^{e_1} \dots p_k^{e_k} \leq \beta \\ e_j > 0 \\ p_j < y}} \Phi\left(\frac{x}{p_1^{e_1} \dots p_k^{e_k}}, y\right) + \sum_{\substack{\beta < p_1^{e_1} \dots p_k^{e_k} \leq x \\ e_j > 0 \\ p_j < y}} 1,$$

and utilizing the above lemmas.

When $x \rightarrow \infty$, if we now permit β to become arbitrarily large (but keep $y > \sqrt{x}$), then the first term in Theorem 2 will begin to dominate, which is the phenomenon noticed in 2015.

The above analysis will permit us to estimate $N_k(x, y)$ for all $y \in (\sqrt{x}, x]$, provided k is fixed, because the only fact used in the course of the proof of Theorem 2 is that if $x > x/n \geq y > \sqrt{x}$ then $\Phi(x/n, y) = \pi(x/n) - \pi(y - 1)$. In addition, the sum

$$\begin{aligned} \sum_{p \leq \beta} \pi\left(\frac{x}{p}\right) &\sim \sum_{p \leq \beta} \frac{x}{p \log(x/p)} = \frac{x}{\log x} \sum_{p \leq \beta} \frac{1}{p\left(1 - \frac{\log p}{\log x}\right)} \\ &= \frac{x}{\log x} \left(\sum_{p \leq \beta} \frac{1}{p}\right) + O\left(\frac{x}{\log^2 x} \sum_{p \leq \beta} \frac{\log p}{p}\right) \\ &= \frac{x}{\log x} \log \log \beta + O\left(\frac{x}{\log x}\right). \end{aligned}$$

This analysis can be continued when we have k prime factors as

$$\begin{aligned} \sum_{p_1 \dots p_k \leq \beta} \pi\left(\frac{x}{p_1 \dots p_k}\right) &= \frac{x}{\log x} \sum_{p_1 \dots p_k \leq \beta} \frac{1}{p_1 \dots p_k} \\ &+ O\left(\frac{x}{\log^2 x} \sum_{p_1 \dots p_k} \frac{\log(p_1 \dots p_k)}{p_1 \dots p_k} + \sum_{p_1 \dots p_k \leq \beta} \frac{x}{p_1 \dots p_k \log^2(x/p_1 \dots p_k)}\right), \end{aligned}$$

so that, with the aid of Lemma 2, we may arrive at the following

Theorem 3 For $\sqrt{x} < y \leq x$, with $\beta = x/y$, we have

$$\begin{aligned} N_k(x, y) &= \frac{x}{\log x} \frac{(\log \log \beta^*)^k}{k!} + \frac{x}{\log x} \frac{(\log \log x)^{k-1}}{(k-1)!} \\ &+ O\left(\frac{x(\log \log x)^{k-2}}{\log x (k-2)!} + \frac{x}{\log x} \frac{(\log \log \beta^*)^{k-1}}{(k-1)!}\right), \end{aligned}$$

where $\beta^* = \max(\beta, 10)$.

We may rewrite Theorem 3 as

$$N_k(x, y) = w(\alpha) \frac{x}{\log y} \frac{(\log \log \beta^*)^k}{k!} + \frac{x}{\log x} \frac{(\log \log x)^{k-1}}{(k-1)!} + O\left(\frac{x(\log \log x)^{k-2}}{\log x (k-2)!} + \frac{x}{\log x} \frac{(\log \log \beta^*)^{k-1}}{(k-1)!}\right), \tag{11}$$

where $w(\alpha) = \frac{1}{\alpha}$ for $1 \leq \alpha < 2$.

We note the following interesting consequence of the preceding theorem.

Corollary 2 *Let $k \in \mathbb{Z}^+$ be fixed and $\sqrt{x} < y < xe^{-e^{(1/\epsilon)(\log \log x)^{1-1/k}}}$, where $\epsilon \rightarrow 0$, as $x \rightarrow \infty$. Then*

$$N_k(x, y) \sim w(\alpha) \frac{x(\log \log \beta)^k}{k! \log y}.$$

Proof The corollary follows by comparing the first two terms of Eq. (11), and noting that for the stated range the first term dominates. □

It is also possible to show that the results of Corollary 2 hold for arbitrary fixed $\alpha > 2$ and k fixed. For $y < \sqrt{x}$, we have for $k = 0$

$$N_0(x, y) = \Phi(x, y) = w(\alpha) \frac{x}{\log y} + O\left(\frac{x}{\log^2 y}\right),$$

where

$$w(\alpha) = \begin{cases} \frac{1}{\alpha}, & \text{for } 1 \leq \alpha \leq 2 \\ \frac{d}{d\alpha} (\alpha w(\alpha)) = w(\alpha - 1), & \text{for } \alpha \geq 2. \end{cases}$$

This is a well-known result (see [5]) for the number of uncanceled elements in the sieve of Eratosthenes, and $w(\alpha)$ is known as the Buchstab function.

Let $\wp_y = \prod_{p < y} p$. When $k = 1$, we see that

$$\begin{aligned} N_1(x, y) &= \sum_{\substack{n=mp^e \leq x \\ (m, \wp_y)=1 \\ p < y}} 1 = \sum_{\substack{p < y \\ e \geq 1}} \Phi\left(\frac{x}{p^e}, y\right) \\ &= \sum_{p < y} \Phi\left(\frac{x}{p}, y\right) + \sum_{\substack{p < y \\ e \geq 2}} \Phi\left(\frac{x}{p^e}, y\right) = \sum_1 + \sum_2. \end{aligned} \tag{12}$$

Clearly,

$$\sum_2 = \sum_{\substack{p^e \leq x/y \\ p < y \\ e \geq 2}} \Phi\left(\frac{x}{p^e}, y\right) + \sum_{\substack{p^e > x/y \\ p < y, p^e \leq x \\ e \geq 2}} \Phi\left(\frac{x}{p^e}, y\right). \tag{13}$$

Using the estimate

$$\Phi(x, y) \ll \max\left(\frac{x}{\log y}, 1\right),$$

we see that (13) is

$$\ll \sum_{\substack{p < y \\ p^e \leq x \\ e \geq 2}} \max\left(\frac{x}{p^e \log y}, 1\right) \ll \frac{x}{\log y} + \sqrt{x} \ll \frac{x}{\log y}. \tag{14}$$

Regarding the first sum in (12), we use the fact that for $0 \leq u \leq 1$

$$w(\alpha - u) = w(\alpha) + O(u),$$

because $w'(\alpha)$ is bounded for $\alpha > 1$ (see either [2] or [5]). Thus

$$\begin{aligned} \sum_1 &= \sum_{p < y} \frac{x}{p} \frac{w\left(\frac{\log x - \log p}{\log y}\right)}{\log y} + O\left(\frac{x}{\log^2 y} \sum_{p < y} \frac{1}{p}\right) \\ &= w(\alpha) \frac{x}{\log y} \sum_{p < y} \frac{1}{p} + O\left(\frac{x}{\log^2 y} \sum_{p < y} \frac{\log(p)}{p}\right) \\ &= w(\alpha) \frac{x \log \log y}{\log y} + O\left(\frac{x}{\log y}\right). \end{aligned} \tag{15}$$

Using (14) and (15) in (12) we conclude

$$N_1(x, y) = w(\alpha) \frac{x \log \log y}{\log y} + O\left(\frac{x}{\log y}\right).$$

Similarly, we may also estimate $N_2(x, y)$ for fixed $\alpha > 2$, but this is a bit more complicated. Note that

$$\begin{aligned} N_2(x, y) &= \sum_{\substack{n = mp_1^{e_1} p_2^{e_2} \leq x \\ (m, \varphi_y) = 1 \\ p_1, p_2 < y}} 1 = \sum_{p_1, p_2 < y} \Phi\left(\frac{x}{p_1^{e_1} p_2^{e_2}}, y\right) \\ &= \sum_{p_1, p_2 < y} \Phi\left(\frac{x}{p_1 p_2}, y\right) + \sum_{\substack{p_1, p_2 < y \\ e_1 + e_2 \geq 3}} \Phi\left(\frac{x}{p_1^{e_1} p_2^{e_2}}\right) = \sum_3 + \sum_4. \end{aligned} \tag{16}$$

In this case

$$\sum_4 \ll \sum_{\substack{p_1, p_2 < y \\ x/y \leq p_1^{e_1} p_2^{e_2} \leq x \\ e_1 + e_2 \geq 3}} \max\left(\frac{x}{p_1^{e_1} p_2^{e_2} \log y}, 1\right) \ll \frac{x \log \log y}{\log y}.$$

With regard to \sum_3 , we have

$$\begin{aligned} \sum_3 &= \sum_{p_1, p_2 < y^{1/4}} \Phi\left(\frac{x}{p_1 p_2}, y\right) + \sum_{\substack{p_1, p_2 < y \\ \max(p_1, p_2) > y^{1/4}}} \Phi\left(\frac{x}{p_1 p_2}, y\right) \\ &= \sum_5 + \sum_6. \end{aligned} \tag{17}$$

Now, the latter sum in (17) is

$$\sum_6 \ll \sum_{\substack{p_1, p_2 < y \\ \text{one of } p_1, p_2 > y^{1/4}}} \frac{x}{p_1 p_2 \log y} + \sum_{\substack{p_1, p_2 < y \\ \max(p_1, p_2) > y^{1/4} \\ x/y < p_1 p_2 < x}} 1, \tag{18}$$

and the first sum in (18) may be bounded by

$$\begin{aligned} \sum_{\substack{p_1, p_2 < y \\ \text{one of } p_1, p_2 > y^{1/4}}} \frac{x}{p_1 p_2 \log y} &\ll \frac{x}{\log y} \left(\sum_{y^{1/4} < p < y} \frac{1}{p} \right) \left(\sum_{q < y} \frac{1}{q} \right) \\ &\ll \frac{x \log \log y}{\log y}. \end{aligned}$$

To bound the second sum in (18) note that if $p_1 p_2 > x/y = x^{1-1/\alpha}$, $\alpha > 2$, then $\max(p_1, p_2) > y^{1/4}$ and so

$$\sum_{\substack{p_1, p_2 < y \\ \max(p_1, p_2) > x^{1/4} \\ x/y < p_1 p_2 < x}} 1 \ll \sum_{x^{1/4} < p_1 < y} \pi\left(\frac{x}{p_1}\right) \ll \frac{x}{\log y} \sum_{x^{1/4} < p_1 < x} \frac{1}{p_1} \ll \frac{x}{\log y}.$$

Lastly,

$$\begin{aligned} \sum_5 &= \sum_{p_1, p_2 < y^{1/t}} \frac{x}{\log y} \frac{w\left(\frac{\log x - \log(p_1 p_2)}{\log y}\right)}{p_1 p_2} + O\left(\frac{x \log \log y}{\log^2(y)}\right) \\ &= w(\alpha) \frac{x}{\log y} \sum_{p_1, p_2 < y^{1/t}} \frac{1}{p_1 p_2} + O\left(\frac{x}{\log^2 y} \sum_{p_1, p_2 < y^{1/4}} \frac{\log(p_1 p_2)}{p_1 p_2}\right) \end{aligned}$$

$$= w(\alpha) \frac{x}{\log y} \frac{(\log \log(y^{1/4}))^2}{2} + O\left(\frac{x \log \log y}{\log y}\right) \tag{19}$$

so that from (16), (17), (18), and (19), we have

$$N_2(x, y) = w(\alpha) \frac{x(\log \log y)^2}{2 \log y} + O\left(\frac{x \log \log y}{\log y}\right).$$

In the above equation it does not matter whether we have $\log x$ or $\log y$ in the denominator of the error term, but we have preferred $\log y$ for reasons which will become clear in the sequel.

This method can be used for $k > 2$, the major difference in this case being that the primes in the above sum which were truncated at $y^{1/4}$ must now be truncated at $y^{1/t}$ with $t = 2k$. The method then yields

Theorem 3* For fixed k and fixed $\alpha > 2$, we have

$$N_k(x, y) = w(\alpha) \frac{x}{\log y} \frac{(\log \log y)^k}{k!} + O\left(\frac{x(\log \log y)^{k-1}}{(k-1)! \log y} \log k\right).$$

Remark In Theorem 3* the $\log k$ factor in the error term is due to

$$\sum_{y^{1/2k} \leq p \leq y} \frac{1}{p} \asymp \log k.$$

We will see in Sect. 6 that by analytic methods, we can derive an estimate of $N_k(x, y)$ for fixed $\alpha > 2$ which is superior to Theorem 3*.

4 Estimate of $S_z(x, y)$ for large y , for $Re(z) > 0$

This section will focus on estimates for $S_z(x, y)$ for large values of y . By large we mean values of y such that $y \geq \exp(\log^{1-\epsilon} x)$ for any sufficiently small $\epsilon > 0$. Our asymptotic estimates for these large values of y are valid only when $Re(z) > 0$. Nevertheless, a result of Tenenbaum [15] (discussed later) ensures that this is all the information that we need.

Let us begin with the case $y \geq x$. Let $R > 0$ be arbitrary but fixed. For $|z| \leq R$, with $r = Re(z)$, we have

$$S_z(x, y) = S_z(x, x) = \frac{g(1, z)}{\Gamma(z)} \frac{x}{\log^{1-z} x} + O\left(\frac{x}{\log^{2-r} x}\right) \tag{20}$$

by the work of Selberg [12]. We shall see that the limit

$$m_z(\alpha) := \lim_{x \rightarrow \infty} \frac{S_z(x, y)}{x / \log^{1-z} y}$$

exists for $\alpha > 1$. However, for $0 < \alpha \leq 1$ we may rewrite equation (20) as

$$S_z(x, y) = m_z(\alpha) \frac{x}{\log^{1-z} y} + O_\alpha \left(\frac{x}{\log^{2-r} y} \right),$$

where

$$m_z(\alpha) = \frac{g(1, z)}{\alpha^{1-z} \Gamma(z)}.$$

With the aid of the following lemma we will be able to show that the function $m_z(\alpha)$ does in fact exist for larger ranges of α .

Lemma 5 For $x \geq y^h \geq y \geq 3$ and $|z| \leq R$, we have

$$S_z(x, y) = S_z(x, y^h) + (1 - z) \int_y^x S_z(x/t, t) \frac{dt}{\log t} + O(xE(y) \log^{R+2} x),$$

where $E(y)$ bounds from above the relative error in the Prime Number Theorem.

Proof From the generalization of the sieve to strongly multiplicative functions $g(n)$ as in [4], we may choose $g(n) = z^{\omega_y(n)}$ to conclude that

$$S_z(x, y) = S_z(x, y^h) + (1 - z) \sum_{y \leq p \leq y^h} S_z(x/p, p). \tag{21}$$

From the above equation we see that the lemma will be proven provided

$$\sum_{y \leq p \leq y^h} S_z(x/p, p) = \int_y^{y^h} S_z(x/t, t) \frac{dt}{\log(t)} + O(xE(y) \log^{R+2}(x)).$$

To establish this, we decompose the intervals in a certain way. When $p \leq n$ and $t \leq n$ we estimate the sum by decomposing the interval $[1, n]$ into the subintervals $I_1 = [1, p_1], I_2 = (p_1, p_2], \dots, I_{\omega(n)} = (p_{\omega(n)-1}, p_{\omega(n)}], I_{\omega(n)+1} = (p_{\omega(n)}, n]$, where $p_1 < p_2 < \dots < p_{\omega(n)}$ denote the distinct prime divisors of n . If $\chi_j(t)$ denotes the characteristic function of the interval I_j , then in this case the sum in (21) equals

$$\begin{aligned} &= \sum_{n \leq x/y} \sum_{j=1}^{\omega(n)+1} \left(\sum_{y < p < \min(x/n, y^h)} \chi_j(n) z^{\omega_p(n)} - \int_y^{\min(x/n, y^h)} \chi_j(t) z^{\omega_t(n)} \frac{dt}{\log t} \right) \\ &= \sum_{n \leq x/y} \sum_{j=1}^{\omega(n)+1} z^{\omega_{p_j}(n)} \left(\sum_{y < p < \min(x/n, y^h)} 1 - \int_y^{\min(x/n, y^h)} \frac{dt}{\log t} \right) \end{aligned}$$

$$\begin{aligned} &<< \sum_{n \leq x/n} \sum_{j=1}^{\omega(n)+1} (1 + [R])^{\omega(n)} \frac{x}{n} E(y) = x E(y) \sum_{n \leq x/y} \frac{(1 + [R])^{\omega(n)}}{n} (\omega(n) + 1) \\ &<< x E(y) \log x \sum_{n \leq x/y} \frac{(1 + [R])^{\omega(n)}}{n} << x E(y) \log^{R+2} x. \end{aligned} \tag{22}$$

The result then follows from (21) and (22). □

Remark We will apply Lemma 5 only when $\log y \gg \log x$ and so by our assumptions in Sect. 1, the expression on the right will be $o(x/\log^U y)$, for every $U > 0$.

We now derive an estimate for the series $S_z(x, y)$ with $1 \leq \alpha \leq 2$ using (21):

Theorem 4 For $2 \leq \sqrt{x} \leq y \leq x$, and $z \in \mathbb{C}$, $|z| \leq R$, $r = \text{Re}(z) > 0$, we have

$$S_z(x, y) = m_z(\alpha) \frac{x}{\log^{1-z} y} + O\left(\frac{|1-z|x \log \log x}{\log^{2-r} y} + \frac{|1-z|x \log \log x}{\log x}\right),$$

and

$$m_z(\alpha) = \frac{g(1, z)}{\Gamma(z)} \left(\frac{1}{\alpha^{1-z}} + \frac{(1-z)}{\alpha^{1-z}} \int_1^\alpha \frac{du}{u^z(u-1)^{1-z}} \right) \tag{23}$$

for $1 \leq \alpha \leq 2$.

Proof We begin with (21) and take $y^h = x$ to obtain

$$S_z(x, y) = S_z(x, x) + (1-z) \sum_{y \leq p < x} S_z(x/p, p). \tag{24}$$

For $S_z(x, x) = S_z(x)$, we state the Selberg estimate for $S_z(x)$ in the form

$$S_z(x) = \frac{g(1, z)}{\Gamma(z)} \frac{x}{\max(\log x, 2)^{1-z}} + O\left(\frac{x}{\max(\log x, 2)^{2-r}}\right). \tag{25}$$

Now, if $\sqrt{x} \leq y \leq x$ and $y \leq p < x$, then $x/p \leq \sqrt{x} \leq y \leq p$ so that the sum in (24) becomes

$$\sum_{y \leq p < x} S_z(x/p, p) = \sum_{y \leq p < x} S_z(x/p) = \int_y^x S_z(x/t) \frac{dt}{\log t} + O(x E(y) \log^{R+2} x)$$

by the results in Lemma 6. It therefore suffices to obtain an accurate estimate of the above integral, which can be done by applying Lemma 6 in the following form:

$$\int_y^x S_z(x/t) \frac{dt}{\log t} = \frac{g(1, z)}{\Gamma(z)} \int_y^{x/e^2} \frac{x}{t \log^{1-z}(x/t) \log t} dt + O\left(\int_y^{x/e^2} \frac{x}{t \log^{2-r}(x/t) \log t} dt + \int_{x/e^2}^x \frac{x}{t \log t} dt\right). \tag{26}$$

If we put $t = x^{1/u}$, the first integral in (26) becomes

$$\int_y^{x/e^2} \frac{x}{t \log^{1-z}(x/t) \log t} dt = \frac{x}{\log^{1-z} x} \int_{\left(1-\frac{2}{\log x}\right)^{-1}}^\alpha \frac{du}{u^z(u-1)^{1-z}}. \tag{27}$$

Also, the second integral in the O -term in (26) is

$$\int_{x/e^2}^x \frac{x}{t \log t} dt \ll \frac{x}{\log x}. \tag{28}$$

We see from (26), (27), and (28) that

$$\int_y^x S_z(x/t) \frac{dt}{\log t} = \frac{g(1, z)}{\Gamma(z)} \frac{x}{\log^{1-z} x} \int_{\left(1-\frac{2}{\log x}\right)^{-1}}^\alpha \frac{du}{u^z(u-1)^{1-z}} + O\left(I_2 + \frac{x}{\log x}\right), \tag{29}$$

where

$$I_2 := \int_y^{x/e^2} \frac{x}{t \log^{2-r} x/t \log t} dt.$$

We write the integral in the main term of Eq.(29) as

$$\int_{\left(1-\frac{2}{\log x}\right)^{-1}}^\alpha \frac{du}{u^z(u-1)^{1-z}} = \int_1^\alpha \frac{du}{u^z(u-1)^{1-z}} - \int_1^{\left(1-\frac{2}{\log x}\right)^{-1}} \frac{du}{u^z(u-1)^{1-z}}. \tag{30}$$

Note that $1 \leq \alpha \leq 2$ so that if $Re(z) > 0$, then

$$\left| \int_1^\alpha \frac{(u-1)^{z-1}}{u^z} du \right| < \infty.$$

Here is where we must use the fact that $Re(z) > 0$, as the first integral on the right in (30) would not converge if $\Re(z) \leq 0$.

Next we wish to bound the second integral on the right in (30). To that end

$$\left| \int_1^{\left(1-\frac{2}{\log x}\right)^{-1}} \frac{du}{u^z(u-1)^{1-z}} \right| \ll \int_1^{\left(1-\frac{2}{\log x}\right)^{-1}} \frac{du}{(u-1)^{1-r}} \ll \frac{1}{\log^r x};$$

and consequently

$$\int_{\left(1-\frac{2}{\log x}\right)^{-1}}^\alpha \frac{du}{u^z(u-1)^{1-z}} = \int_1^\alpha \frac{du}{u^z(u-1)^{1-z}} + O\left(\frac{1}{\log^r x}\right). \tag{31}$$

The evaluation of the integral I_2 is similar to what we have done for the main term, except here we must show care when $r = \operatorname{Re}(z)$ is close to 1. This is achieved by considering the two cases: $|\operatorname{Re}(z) - 1| > \frac{1}{5\log \log x}$ and $|\Re(z) - 1| \leq \frac{1}{5\log \log x}$. The substitution $t = x^{1/u}$ shows that

$$I_2 = O\left(\frac{x \log \log x}{\log^{2-r}(x)}\right)$$

for $\operatorname{Re}(z) \geq 1$ and

$$I_2 = O\left(\frac{x \log \log x}{\log x}\right)$$

when $\Re(z) < 1$. Thus

$$\begin{aligned} & \int_y^x S_z(x/t) \frac{dt}{\log t} \\ &= \frac{g(1, z)}{\Gamma(z)} \frac{x}{\log^{1-z} x} \int_1^\alpha \frac{du}{u^z(u-1)^{1-z}} + O\left(\frac{x \log \log x}{\log x} + \frac{x \log \log x}{\log^{2-r}(x)}\right), \end{aligned}$$

from (31), and so the result follows.

We may also note that if $r \in \mathbb{R}^+$ then the preceding result can be stated without the $\log \log x$, that is

Theorem 4* For $2 \leq \sqrt{x} \leq y \leq x$, and $z = r > 0$, we have

$$S_r(x, y) = m_r(\alpha) \frac{x}{\log^{1-r}(y)} + O\left(\frac{x}{\log^{2-r}(y)} + \frac{x}{\log(x)}\right),$$

where $m_r(\alpha)$ is given by (23).

We may now use Buchstab's recurrence (24) to derive an estimate for $S_z(x, y)$ for $\alpha \geq 2$ by applying induction on $[\alpha]$. Fortunately, we will also deduce an asymptotic estimate for $S_z(x, y)$ when $\operatorname{Re}(z) > 0$ for y outside the range of Corollary 1. First we prove

Theorem 5 Let $|z| \leq R$, and $r = \text{Re}(z) > 0$. Then for arbitrary but fixed $\alpha > 2$, we have

$$S_z(x, y) = \frac{x m_z(\alpha)}{\log^{1-z} y} + O\left(\frac{x \log \log x}{\log^{2-r} y} + \frac{x \log \log x}{\log y}\right),$$

where for $\alpha > 2$, $m_z(\alpha)$ is given by

$$m_z(\alpha) = \frac{2^{1-z} m_z(2)}{\alpha^{1-z}} + \frac{1-z}{\alpha^{1-z}} \int_2^\alpha \frac{m_z(u-1)}{u^z} du. \tag{32}$$

Proof Setting $y^h = \sqrt{x}$ in (21), we have

$$\begin{aligned} S_z(x, y) &= S_z(x, \sqrt{x}) + (1-z) \sum_{y \leq p \leq \sqrt{x}} S_z(x/p, p) \\ &= \frac{2^{1-z} m_z(2)}{\alpha^{1-z}} \frac{x}{\log^{1-z} y} + O\left(\frac{|1-z|x \log \log x}{\log^{2-r}(y)}\right) \\ &\quad + (1-z) \int_y^{\sqrt{x}} m_z\left(\frac{\log(x/t)}{\log t}\right) \frac{x/t}{\log^{1-z} t \log t} dt \\ &\quad + O\left(|1-z| \int_y^{\sqrt{x}} \frac{x \log \log x/t}{t \log^{2-r} t \log t} dt\right) \\ &\quad + O\left(|1-z| \int_y^{\sqrt{x}} \frac{x \log \log x}{t \log^2 t} dt\right) \\ &\quad + O(xE(y) \log^R(x)), \end{aligned}$$

as $\sqrt{x} = y^{\alpha/2}$. With the familiar substitution of $t = x^{1/u}$ we obtain from the above

$$\begin{aligned} S_z(x, y) &= \frac{2^{1-z} m_z(2)}{\alpha^{1-z}} \frac{x}{\log^{1-z} y} + O\left(\frac{x \log \log x}{\log^{2-r} y}\right) \\ &\quad + \frac{(1-z)}{\alpha^{1-z}} \frac{x}{\log^{1-z} y} \int_2^\alpha m_z(u-1) \frac{du}{u^z} + O\left(\frac{x \log \log x}{\log^{2-r} y}\right) \\ &\quad + O\left(\frac{x \log \log x}{\log y}\right). \end{aligned}$$

Hence the theorem follows from this with $m_z(\alpha)$ as defined in (32) above.

From the definition of $m_z(\alpha)$ in (32), we see that

$$\alpha^{1-z} m_z(\alpha) - (\alpha-1)^{1-z} m_z(\alpha-1) = (1-z) \int_{\alpha-1}^\alpha \frac{m_z(u-1)}{u^z} du,$$

for $\alpha \geq 3$. This can be rewritten as

$$m_z(\alpha) := \frac{m_z(\alpha - 1)(\alpha - 1)^{1-z}}{\alpha^{1-z}} + \frac{(1 - z)}{\alpha^{1-z}} \int_{\alpha-1}^{\alpha} m_z(u - 1) \frac{du}{u^z}.$$

We next derive an improvement of Theorem 5 in which the asymptotic estimate will hold for α tending to infinity with x . □

Theorem 6 *Let $\alpha \geq 1$, and $|z| \leq R$. If $r = \text{Re}(z) \geq 1$, then there exists an absolute constant $K = K(R)$ such that*

$$\left| S_z(x, y) - m_z(\alpha) \frac{x}{\log^{1-z} y} \right| \ll \frac{\alpha^K x \log \log x}{\log^{2-r} y},$$

and if $0 < \delta \leq r = \text{Re}(z) < 1$, then

$$\left| S_z(x, y) - m_z(\alpha) \frac{x}{\log^{1-z} y} \right| \ll \frac{\alpha^K x \log \log x}{\log y}.$$

Proof It suffices to prove the theorem for $\alpha > 3$, in which case we can use (32). Let $y = x^{1/\alpha}$, $y^h = x^{1/(\alpha-1)}$, and $u = \frac{\log x}{\log t}$. Lemma 6 shows that, with this notation,

$$\begin{aligned} S_z(x, y) &= S_z(x, y^h) + (1 - z) \int_y^{y^h} S_z(x/t, t) \frac{dt}{\log t} + O\left(x\alpha^{|z|+1} E(y) \log^{R+2}(y)\right). \end{aligned} \tag{33}$$

Assume that $\text{Re}(z) \geq 1$. We shall prove the result by induction on α . Assume that there exists a positive, non-decreasing function $\phi(u)$ such that for all $u \leq \alpha - 1$ and $x > y > 1$ we have

$$\left| S_z(x, t) - m_z(u) \frac{x}{\log^{1-z}(t)} \right| < \frac{\phi(u)x \log \log x}{\log^{2-r} t}. \tag{34}$$

Theorem 5 establishes the validity of (34) when $\alpha \in [2, 3]$. By the inductive hypothesis of (34) and (33) we now obtain

$$\begin{aligned} S_z(x, y) &= m_z(\alpha - 1) \left(\frac{\alpha - 1}{\alpha}\right)^{1-z} \frac{x}{\log^{1-z} y} \\ &+ O_1 \left(\left(\frac{\alpha - 1}{\alpha}\right)^{2-r} \frac{\phi(\alpha - 1)x \log \log x}{\log^{2-r} y} \right) \\ &+ (1 - z) \int_y^{y^h} m_z \left(\frac{\log x - \log t}{\log t} \right) \frac{x}{t \log^{2-z} t} dt \end{aligned}$$

$$\begin{aligned}
 &+ O_1 \left(|1 - z| \int_y^{y^h} \phi \left(\frac{\log x - \log t}{\log t} \right) \frac{x \log \log x}{t \log^{3-r} t} dt \right) \\
 &+ O(xE(y)\alpha^{|z|} \log^{R+2} y),
 \end{aligned} \tag{35}$$

where the notation O_1 implies that the implicit constant is ≤ 1 (which will be important when iterating the process). We will see that the first and third terms in the above recurrence will make the largest contribution. Note that with $t = x^{1/u}$

$$\begin{aligned}
 m_z(\alpha - 1) \left(\frac{\alpha - 1}{\alpha} \right)^{1-z} \frac{x}{\log^{1-z} y} &+ (1 - z) \int_y^{y^h} m_z \left(\frac{\log x - \log t}{\log t} \right) \frac{x}{t \log^{2-z} t} \\
 &= m_z(\alpha - 1) \left(\frac{\alpha - 1}{\alpha} \right)^{1-z} \frac{x}{\log^{1-z} y} + \frac{(1 - z)}{\alpha^{1-z}} \frac{x}{\log^{1-z} y} \int_{\alpha-1}^{\alpha} m_z(u - 1) \frac{du}{u^z} \\
 &= m_z(\alpha) \frac{x}{\log^{1-z} y},
 \end{aligned} \tag{36}$$

from (32). We conclude from (35) and (36) that

$$\begin{aligned}
 S_z(x, y) &= \frac{m_z(\alpha)x}{\log^{1-z} y} + O_1 \left(\left(\frac{\alpha - 1}{\alpha} \right)^{2-r} \frac{\phi(\alpha - 1)x \log \log x}{\log^{2-r} y} \right) \\
 &+ O_1 \left(|1 - z|x \log \log x \phi(\alpha - 1) \int_y^{y^h} \frac{dt}{t \log^{3-r} t} \right) \\
 &+ O(x\alpha^{|z|+1} E(y) \log^{R+2} y).
 \end{aligned} \tag{37}$$

If $Re(z) \neq 2$ the integral in (37) will be equal to

$$\begin{aligned}
 \int_y^{y^h} \frac{dt}{t \log^{3-r} t} &= \frac{\log^{r-2}(y^h)}{r - 2} - \frac{\log^{r-2} y}{r - 2} \\
 &= \frac{1}{r - 2} \frac{1}{\log^{2-r} y} \left(1 - \frac{1}{h^{2-r}} \right),
 \end{aligned}$$

and

$$1 - \frac{1}{h^{2-r}} = 1 - \left(1 - \frac{1}{\alpha} \right)^{2-r} = O \left(\frac{|2 - r|}{\alpha} \right). \tag{38}$$

If $\Re(z) = 2$, then the integral is

$$\begin{aligned}
 \int_y^{y^h} \frac{dt}{t \log^{3-r} t} &= \int_y^{y^h} \frac{dt}{t \log t} = \log h \\
 &= \log \left(\frac{\alpha}{\alpha - 1} \right) = 1 + O \left(\frac{1}{\alpha} \right).
 \end{aligned} \tag{39}$$

From Eqs. (37), (38), and (39) Eq. (35) becomes

$$S_z(x, y) = \frac{m_z(\alpha)x}{\log^{1-z} y} + O_1 \left(\left(\frac{\alpha - 1}{\alpha} \right)^{2-r} \frac{\phi(\alpha - 1)x \log \log x}{\log^{2-r} y} \right) + O_1 \left(\frac{|1 - z|x \log \log x \phi(\alpha - 1) C}{\log^{2-r} y \alpha} \right). \tag{40}$$

We are free to choose our function $\phi(u)$ to satisfy

$$\phi(\alpha) < \phi(\alpha - 1) \left(\frac{\alpha - 1}{\alpha} \right)^{1-r} + \phi(\alpha - 1) \frac{|1 - z|C}{\alpha} \leq \phi(\alpha - 1) \left(1 + \frac{|1 - z|C'}{\alpha} \right).$$

□

Because the function $\phi(\alpha)$ is defined recursively we may estimate its growth rate by noting that for any $j \in \mathbb{Z}, 0 \leq j \leq \alpha - 2$

$$\frac{\phi(\alpha - j)}{\phi(\alpha - j - 1)} < 1 + \frac{|1 - z|C}{\alpha - j};$$

hence,

$$\begin{aligned} \phi(\alpha) &= \frac{\phi(\alpha)}{\phi(\alpha - 1)} \frac{\phi(\alpha - 1)}{\phi(\alpha - 2)} \dots \frac{\phi(2)}{\phi(1)} < \prod_{0 \leq j \leq \alpha - 1} \left(1 + \frac{|1 - z|C}{\alpha - j} \right) \\ &= \prod_{1 \leq j \leq \alpha} \left(1 + \frac{|1 - z|C}{j} \right) \ll \alpha^{C|1-z|}. \end{aligned}$$

Collecting the above results, we conclude that

$$\left| S_z(x, y) - m_z(\alpha) \frac{x}{\log^{1-z}(y)} \right| < \frac{\alpha^{C|1-z|} x \log \log x}{\log^{2-r}(y)},$$

with the constant $K = K(R) = C|1 - z|$, thereby proving the first statement of Theorem 6.

If $0 < r = Re(z) < 1$ then the method of Alladi in [1] shows that

$$\left| S_z(x, y) - m_z(\alpha) \frac{x}{\log^{1-z}(y)} \right| \leq \frac{\alpha x \log \log x}{\log y}$$

in place of (34). The second statement of the theorem then follows by repeating the above induction procedure.

The preceding theorem gives the desired uniform result we need to analyze the sum $S_z(x, y)$ for large values of y . Let $r = Re(z) \geq \delta > 0$ and consider

$$\log^\delta y = \alpha^K \log \log x = \left(\frac{\log x}{\log y} \right)^K \log \log x \leq \frac{\log^{K+\epsilon} x}{\log^K y}$$

for some $0 < \epsilon < \delta$, so that

$$\log y = \log^{(k+\epsilon)/(k+\delta)} x = \log^{1-\delta'} x,$$

where

$$\delta' = \frac{\delta - \epsilon}{k + \delta}, \quad \text{and} \quad y = \exp(\log^{1-\delta'}(x)).$$

Therefore, we obtain an asymptotic estimate of $S_z(x, y)$ from Theorem 6 when $y \geq e^{\log^{1-\delta'}(x)}$, which is larger than the range specified in Corollary 1.

Let us pause for a moment and reflect on what has just been proven. In Corollary 1 we obtained an asymptotic formula for $S_z(x, y)$ ($|z| \leq R$), provided $y \leq x^{1/(R+D+1+\epsilon) \log \log x}$; however, Theorem 6 gives an asymptotic formula (for $Re(z) > 0, |z| \leq R$) provided $y \geq x^{1/\log^{\delta'} x}$. Observe that for x sufficiently large, we have

$$x^{1/\log^{\delta'} x} \leq x^{1/(R+D+1+\epsilon) \log \log x}$$

so the ranges of Corollary 2 and Theorem 6 overlap. We have by virtue of Theorem 6 and Corollary 2 derived an asymptotic formula for all $y \leq x$ if $R \geq Re(z) > 0$ (note, however, that the results of Corollary 2 are true with only the restriction $|z| \leq R$). This uniform estimate will be one of the main tools utilized in Sect. 6 for the study of the local distribution of small prime factors.

5 Properties of $m_z(\alpha)$

In this section, we shall study the properties of the function $m_z(\alpha)$ arising in Theorem 5. Throughout this section, α is just a positive real number, although in applying our results on $m_z(\alpha)$ to $S_z(x, y)$, we will choose $\alpha = \frac{\log x}{\log y}$. We have already shown that $m_z(\alpha)$ exists and is given by (32). Thus

$$\alpha^{1-z} m_z(\alpha) = 2^{1-z} m_z(2) + (1 - z) \int_2^\alpha m_z(u - 1) \frac{du}{u^z};$$

so that $m_z(\alpha)$ satisfies the following difference-differential equation:

$$(\alpha^{1-z} m_z(\alpha))' = (1 - z) \left(\frac{m_z(\alpha - 1)}{\alpha^z} \right), \tag{41}$$

where $m'_z(\alpha) = \frac{d}{d\alpha} m_z(\alpha)$. Equation (41) implies that

$$\alpha^{1-z} m'_z(\alpha) + (1 - z) \alpha^{-z} m_z(\alpha) = (1 - z) m_z(\alpha - 1) \alpha^{-z}$$

which can be rewritten as

$$m'_z(\alpha) = \frac{1-z}{\alpha} (m_z(\alpha-1) - m_z(\alpha)) = \frac{z-1}{\alpha} \int_{\alpha-1}^{\alpha} m'_z(u) du. \tag{42}$$

The above equation will allow us to derive some useful properties about the convergence of $m_z(\alpha)$ as $\alpha \rightarrow \infty$.

Lemma 6 For $z \in \mathbb{C}$, $Re(z) > 0$, $|z| \leq R$, and $\alpha \geq 2$ the function $m'_z(\alpha)$ is differentiable. Moreover, $m'_z(\alpha) \ll e^{-\alpha \log \alpha + O_R(\alpha)}$.

Proof We may immediately deduce from (42) that $m'_z(\alpha)$ is differentiable. We will first show that Eq. (42) implies that $m'_z(\alpha)$ is uniformly bounded for $\alpha \in \mathbb{R}$, $\alpha \geq 2$.

Suppose not. Then $|m'_z(\alpha)|$ assumes all sufficiently large values by continuity. Define

$$B := \max_{\alpha \leq |z-1|+1} |m'_z(\alpha)|$$

and consider a value $M > B$ assumed by $|m'_z(\alpha)|$. Next, define

$$\alpha_0 := \inf\{\alpha : |m'_z(\alpha)| = M\}$$

so that $|m'_z(\alpha_0)| = M$ and $\alpha_0 > |z-1| + 1$. Now, from (42) we see

$$M = |m'_z(\alpha_0)| \leq \frac{|z-1|}{\alpha_0} \int_{\alpha_0-1}^{\alpha_0} |m'_z(t)| dt < \frac{|z-1|}{\alpha_0} M < M,$$

a contradiction. It follows that

$$\{\alpha : |m'_z(\alpha)| = M\} = \emptyset$$

for each $M > B$, so that $\max |m'_z(\alpha)| \leq B$.

Using the above analysis, and again invoking equation (42), we see that

$$|m'_z(\alpha)| = \frac{|1-z|}{\alpha} \left| \int_{\alpha-1}^{\alpha} m'_z(u) du \right| \leq \frac{|1-z|}{\alpha} \int_{\alpha-1}^{\alpha} |m'_z(u)| du < \frac{B|1-z|}{\alpha}.$$

Consider the function $\sup_{u \in [t, +\infty)} |m'_z(u)|$ which is clearly monotone and non-increasing, so

$$\begin{aligned} |m'_z(\alpha)| &\leq \frac{|1-z|}{\alpha} \int_{\alpha-1}^{\alpha} \sup_{u \in [t, +\infty)} |m'_z(u)| dt \leq \frac{|1-z|}{\alpha} \sup_{t \in [\alpha-1, +\infty)} |m'_z(t)| \\ &< \frac{|1-z|}{\alpha} \left(\frac{|1-z|}{\alpha-1} B \right) = \frac{|1-z|^2}{\alpha(\alpha-1)} B \end{aligned}$$

and by iteration

$$|m'_z(\alpha)| \leq \frac{|1 - z|^{[\alpha]}}{\Gamma([\alpha] + 1)} B. \tag{43}$$

Since $\frac{1}{\Gamma(\alpha+1)} \ll e^{-\alpha \log \alpha}$, we deduce from (43) that

$$m'_z(\alpha) \ll e^{-\alpha \log \alpha} (1 + R)^{[\alpha]} = e^{-\alpha \log \alpha} e^{[\alpha] \log(1+R)} = e^{-\alpha \log \alpha + O_R(\alpha)}.$$

The rapid rate of decay of $m'_z(\alpha)$ as $\alpha \rightarrow \infty$ forces the function $m_z(\alpha)$ to approach a limit $\ell(z)$. We have already observed in Sect. 2 that for certain ranges of $\alpha = \frac{\log x}{\log y}$, the estimate for the sum $S_z(x, y)$ is supplied by Theorem 1. The fact that the ranges of our estimates overlap will allow us to obtain a representation for the limiting function $\ell(z)$. □

Theorem 7 *There exists a function $\ell(z)$ such that for $|z| \leq R$,*

$$m_z(\alpha) = \ell(z) + O(e^{-\alpha \log \alpha + O(\alpha)})$$

as $\alpha \rightarrow \infty$, and is given by

$$\ell(z) = e^{(z-1)\gamma} \prod_p \left(1 + \frac{z-1}{p}\right) \left(1 - \frac{1}{p}\right)^{z-1}.$$

Thus $\ell(z)$ given by the product is an analytic function of $z \in \mathbb{C}$, and is nonzero if $z \neq 1$ or $1 - p$, where p denotes a prime number.

Proof We have seen in Lemma 7 that

$$\frac{d}{d\alpha} m_z(\alpha) = m'_z(\alpha) \ll \frac{|z - 1|^\alpha}{\Gamma(\alpha)} \tag{44}$$

so that

$$f(z) := \int_1^\infty m'_z(t) dt \tag{45}$$

exists, because (44) implies the convergence of the integral in (45). Now,

$$\int_1^\alpha m'_z(t) dt = m_z(\alpha) - m_z(1)$$

and so

$$\int_\alpha^\infty m'_z(t) dt = \int_1^\infty m'_z(t) dt - \int_1^\alpha m'_z(t) dt = f(z) - m_z(\alpha) + m_z(1). \tag{46}$$

Thus from (46) and (47) we obtain

$$m_z(\alpha) = f(z) + m_z(1) + O_R(e^{-\alpha \log \alpha + O_R(\alpha)}) \tag{47}$$

thereby proving the limit exists with

$$\ell(z) = f(z) + m_z(1) = \int_1^\infty m'_z(t) dt + m_z(1).$$

□

Now, recall from the comments following the proof of Theorem 6 that if we choose $(R + D + 2) \log \log x < \alpha < (R + D + 3) \log \log x$ then we may apply both Theorem 6 and Corollary 1; therefore, we need only equate $\ell(z)$ with the corresponding term in Corollary 1. For $\alpha \asymp \log \log x$ Corollary 1 supplies

$$\begin{aligned} \frac{S_z(x, y)}{x / \log^{1-z} y} &= \log^{1-z} y \prod_{p < y} \left(1 + \frac{z-1}{p}\right) + O\left(\frac{1}{\log^\epsilon y}\right) \\ &= e^{(z-1)\gamma} \prod_{p < y} \left(1 + \frac{z-1}{p}\right) \left(1 - \frac{1}{p}\right)^{z-1} + O\left(\frac{1}{\log^\epsilon y}\right), \end{aligned}$$

so that by letting as $x, y, \alpha \rightarrow \infty$ we see that

$$\ell(z) = e^{(z-1)\gamma} \prod_p \left(1 + \frac{z-1}{p}\right) \left(1 - \frac{1}{p}\right)^{z-1}.$$

Having shown that the limit $\lim_{\alpha \rightarrow \infty} m_z(\alpha) = \ell(z)$ exists, it is clear from the product formula of $\ell(z)$ that the limit is nonzero for $z \neq 1$ or $1 - p$. However, for future applications we must demonstrate that $m_r(\alpha) > 0$ for all $r > 0$. With this in mind, we now prove the following corollary.

Corollary 3 *The function $m_r(\alpha) > 0$ for all real $r \geq 0$ and $\alpha \geq 1$.*

Proof Let $0 \leq r < 1$ so that we may immediately conclude from (20) that $m_r(1) > 0$. Now we may proceed by induction on $[\alpha]$ and utilize (32) to see that $m_r(\alpha) > 0$ for all $\alpha \geq 1$.

If $r = 1$ then the function $m_1(\alpha) \equiv 1$, so we need only consider the case when r is greater than 1 for the corollary to be proven. However, if $r > 1$ then for a fixed x , $S_r(x, y) = \sum_{n \leq x} r^{\omega_y(n)}$ will be an increasing function of y and, consequently $S_r(x, y)$ will be a decreasing function of α . Let us define the function $m_r^*(\alpha)$ to satisfy

$$S_r(x, y) = m_r^*(\alpha) \frac{x}{\log^{1-r} x} + O\left(\frac{x \log \log x}{\log^{2-r} x} + \frac{x \log \log x}{\log x}\right)$$

and so $m_r^*(\alpha)\alpha^{r-1} = m_r(\alpha)$, with $m_r^*(\alpha)$ decreasing. As $S_r(x, y) > 0$ we see that $m_r^*(\alpha) > 0$ for all $\alpha \geq 1$, and this forces $m_r(\alpha) > 0$ for all $\alpha \geq 1$. \square

Theorem 8 For each $\alpha > 1$

$$\lim_{r \rightarrow 0^+} m_r(\alpha) = w(\alpha).$$

Proof Note that for $r > 0, 1 < \alpha < 2, m_r(\alpha)$ is given by

$$m_r(\alpha) = \frac{s(r)}{\alpha^{1-r}} + \frac{(1-r)s(r)}{\alpha^{1-r}} \int_1^\alpha \frac{du}{u^r(u-1)^{1-r}},$$

where $s(r) = g(1, r)/\Gamma(r)$, and for $\alpha > 2$

$$\left(\alpha^{1-r} m_r(\alpha)\right)' = (1-r) \frac{m_r(\alpha-1)}{\alpha^r}. \tag{48}$$

Clearly when $r \rightarrow 0^+$, the differential equation in (48) coincides with that of the Buchstab function given in Sect. 3. So we only prove that $\lim_{r \rightarrow 0^+} m_r(\alpha) = w(\alpha)$, for $1 < \alpha < 2$. So we must show that

$$\lim_{r \rightarrow 0^+} \left(\frac{s(r)}{\alpha^{1-r}} + \frac{(1-r)s(r)}{\alpha^{1-r}} \int_1^\alpha \frac{du}{u^r(u-1)^{1-r}} \right) = \frac{1}{\alpha}.$$

Since $\lim_{r \rightarrow 0^+} s(r) = 0$, we must show that

$$\lim_{r \rightarrow 0^+} \frac{(1-r)s(r)}{\alpha^{1-r}} \int_1^\alpha \frac{du}{u^r(u-1)^{1-r}} = \frac{1}{\alpha}. \tag{49}$$

Note that $s(r) \sim r$ as $r \rightarrow 0^+$, because $g(1) = 1$ and $\Gamma(r)$ has a simple pole with residue 1 at $r = 0$. Using $u^{-r} = e^{-r \log(u)} = 1 + O(r)$, for $1 < u < \alpha < 2$ and $r \rightarrow 0^+$, we have

$$\begin{aligned} \int_1^\alpha \frac{du}{u^r(u-1)^{1-r}} &= \int_1^\alpha (1 + O(r)) \frac{du}{(u-1)^{1-r}} \\ &= \int_1^\alpha \frac{du}{(u-1)^{1-r}} + O\left(r \int_1^\alpha \frac{du}{(u-1)^{1-r}}\right) = \frac{(\alpha-1)^r}{r} + O((\alpha-1)^r). \end{aligned} \tag{50}$$

From (50) we obtain

$$\frac{(1-r)s(r)}{\alpha^{1-r}} \int_1^\alpha \frac{du}{u^r(u-1)^{1-r}} = \frac{(1-r)s(r)}{\alpha^{1-r}} \left(\frac{(\alpha-1)^r}{r} + O((\alpha-1)^r) \right), \tag{51}$$

and taking the limit as $r \rightarrow 0^+$ in (51) gives the limit in (49) as

$$= \lim_{r \rightarrow 0^+} \frac{(1-r)s(r)}{\alpha^{1-r}} \left(\frac{(\alpha-1)^r}{r} + O((\alpha-1)^r) \right) = \frac{1}{\alpha} \lim_{r \rightarrow 0^+} \frac{s(r)}{r} = \frac{1}{\alpha}.$$

This proves the theorem. □

Remark We had defined $m_z(\alpha)$ only for $Re(z) > 0$, but $S_0(x, y) = \Phi(x, y)$. Thus it is important to establish Theorem 8, from which we could interpret $w(\alpha)$ as $m_0(\alpha)$.

6 The local distribution of the number of small prime factors

In this section, we will apply the analytic results obtained in Sects. 2, 4, and 5 to study the function $N_k(x, y)$ by the contour integral method of Selberg in [12], by which we mean that we will apply the Cauchy integral formula to $S_z(x, y)$ in the following form

$$N_k(x, y) = \frac{1}{2\pi i} \int_{\partial U_\rho} \frac{S_z(x, y)}{z^{k+1}} dz, \tag{52}$$

U_ρ being the circle of radius ρ centered at the origin. We will begin with an analysis of the results which can be obtained from Theorem 1, which holds uniformly for *small* y and $|z| \leq R$. We will then study $N_k(x, y)$ for *large* values of y , in which case our results only apply for $Re(z) > 0$. However, as was alluded to at the end of Sect. 4, Theorem 11 due to Tenenbaum [15] will allow us to estimate $N_k(x, y)$ for *large* values of y provided we have an estimate for $S_r(x, y)$ for real $r > 0$, which will be supplied by the results of Sect. 4.

Tenenbaum [14,17] has also supplied the following alternative representation for the function $m_r(\alpha)$:

Theorem 9 *We have*

$$m_r(\alpha) = C(r) \left(\int_0^{\alpha-1} w(\alpha-t) \rho_r(t) dt + \rho_r(\alpha) \right),$$

where $w(u)$ is the Buchstab function, $\rho_r(u)$ is a function which for $0 < u \leq 1$ equals

$$\rho_r(u) = \frac{u^{r-1}}{\Gamma(r)}$$

and for $u > 1$ satisfies the differential equation

$$u \rho_r'(u) + (1-r) \rho_r(u) + r \rho_r(u-1) = 0,$$

and

$$C(r) = \prod_p \left(1 - \frac{1}{p} \right)^r \left(1 + \frac{r}{p-1} \right).$$

The function ρ_z arises in the asymptotic analysis of

$$\Psi_z(x, y) = \sum_{\substack{n \leq x \\ P^+(n) \leq y}} z^{\omega(n)},$$

where

$$\Psi_z(x, y) \sim \rho_z(\alpha) \frac{x}{\log^{1-z}(y)}.$$

In particular, $\rho_z(\alpha) = e^{-\alpha \log \alpha + O(\alpha)}$. This rapid decay, coupled with the fact that

$$\int_0^\infty \rho_z(\alpha) d\alpha = e^{z\gamma},$$

ensures that $m_r(\alpha) \rightarrow \ell(r)$.

The Selberg method can now be applied to the results of Theorem 1. The consequence is the following theorem, which provides an estimate of $N_k(x, y)$ for small values of y .

Theorem 10 *There exists a constant $C > 0$ such that if $\alpha > C \log \log x$, $k \geq 1$, and $r > 0$, then*

$$N_k(x, y) = \ell\left(\frac{k}{\log \log y}\right) \frac{x}{\log y} \frac{(\log \log y)^k}{k!} \left(1 + O\left(\frac{k}{(\log \log y)^2}\right)\right),$$

uniformly for $k \leq r \log \log y$.

Remarks Our proof is based upon the method of Selberg in [12] (generalized in Chapter II.6.1 of [13]). However, there is an important difference between his approach and the following proof of Theorem 10. In Selberg's estimate for the sum $S_z(x)$, there is a $\frac{1}{\Gamma(z)}$ factor which will, by virtue of the functional equation of the Γ -function, absorb the factor $\frac{1}{z}$ in $\frac{1}{z^{k+1}}$ in Eq. (52). In view of this, Selberg's choice for the radius of the circle in (52), namely $\rho = \frac{k-1}{\log \log x}$, is optimal due to the vanishing of a first-order error term similar to what is given in (59) below. In contrast, our estimate of $S_z(x, y)$ does not contain a factor involving the Γ -function, and since the function $\ell(z)$ in Theorem 7 has the property that $\ell(0) \neq 0$, it does not absorb any of the factors of z^{-k-1} at $z = 0$. We will see in the course of the proof of Theorem 10 that the optimal choice for the radius in (52) will be $\rho = \frac{k}{\log \log y}$ (to ensure the vanishing of the first-order error term in the estimation of $S_z(x, y)$). The absence of the Γ -function is the reason why $N_k(x, y)$ is to be compared with $N_{k+1}(x)$ and as will be seen below.

Proof From Theorem 1, with $|z| \leq R$, we have, for

$$y \leq x^{1/(R+D+2) \log \log x},$$

(D the constant in Theorem 1) equivalently, $\alpha \geq (R + D + 2) \log \log x$

$$S_z(x, y) = x \prod_{p < y} \left(1 + \frac{z-1}{p}\right) + O(xe^{-\alpha} \log^D(x)) + O\left(\frac{x}{\log^{R+2} x}\right).$$

The above equation is then

$$S_z(x, y) = \ell(z) \frac{x}{\log^{1-z} y} + O\left(\frac{x}{\log^{2-r} y}\right)$$

and so

$$\begin{aligned} N_k(x, y) &= \frac{1}{2\pi i} \int_{|z|=\rho} S_z(x, y) \frac{dz}{z^{k+1}} \\ &= \frac{x}{2\pi i \log y} \int_{|z|=\rho} \frac{\ell(z) \log^z y}{z^{k+1}} dz + O\left(\frac{x}{\log^2 y} \int_{|z|=\rho} \frac{\log^{Re(z)} y}{|z|^{k+1}} |dz|\right) \\ &= \frac{x}{\log y} \frac{1}{2\pi i} \int_{|z|=\rho} \frac{\ell(z) \log^z y}{z^{k+1}} dz + O\left(\frac{x}{\log^2 y} \frac{e^{r \log \log y}}{r^{k+1}}\right) \end{aligned} \tag{53}$$

as $Re(z) \leq |z| = \rho$. Upon choosing $\rho = \frac{k}{\log \log y}$, (53) yields

$$\begin{aligned} &= \frac{x}{\log y} \frac{1}{2\pi i} \int_{|z|=\rho} \frac{\ell(z) \log^z y}{z^{k+1}} dz + O\left(\frac{x}{\log^2 y} \frac{e^k}{k^k} (\log \log y)^k\right) \\ &= \frac{x}{\log y} \frac{1}{2\pi i} \int_{|z|=\rho} \frac{\ell(z) \log^z y}{z^{k+1}} dz + O\left(\frac{x}{\log^2 y} \frac{\sqrt{k} (\log \log y)^k}{k!}\right) \end{aligned} \tag{54}$$

from a weak form of Stirling's formula $k! \ll k^{k+1/2} e^{-k}$. □

Recall Theorem 7 where it was noted that $\ell(z)$ is analytic, and so we may write

$$\ell(z) = \ell(\rho) + \ell'(\rho)(z - \rho) + O((z - \rho)^2) \tag{55}$$

so that using (55) we may represent the integral on the right of (54) in the following form

$$\begin{aligned} &\frac{x}{\log y} \frac{1}{2\pi i} \int_{|z|=\rho} \frac{\ell(z) \log^z y}{z^{k+1}} dz \\ &= \frac{x}{\log y} \frac{\ell(\rho)}{2\pi i} \int_{|z|=\rho} \frac{e^{z \log \log y}}{z^{k+1}} dz + \frac{x}{\log y} \frac{\ell'(\rho)}{2\pi i} \int_{|z|=\rho} (z - \rho) \frac{e^{z \log \log y}}{z^{k+1}} dz \\ &\quad + O\left(\frac{x}{\log y} \int_{|z|=\rho} (z - \rho)^2 \frac{\log^{Re(z)} y}{z^{k+1}} dz\right) \\ &= I_1 + I_2 + I_3. \end{aligned}$$

The integral I_1 may be easily evaluated by direct appeal to the Cauchy integral formula:

$$I_1 = \frac{x}{\log y} \frac{\ell(\rho)}{2\pi i} \int_{|z|=\rho} \frac{e^{z \log \log y}}{z^{k+1}} dz = \ell \left(\frac{k}{\log \log y} \right) \frac{x}{\log y} \frac{(\log \log y)^k}{k!} \tag{56}$$

upon identifying the integral with the coefficient of z^k in the series expansion of $e^{z \log \log y}$. With regard to the integral I_2 , note that

$$\begin{aligned} & \frac{\ell'(\rho)}{2\pi i} \int_{|z|=\rho} (z - \rho) \frac{e^{z \log \log y}}{z^{k+1}} dz \\ &= \frac{\ell'(\rho)}{2\pi i} \int_{|z|=\rho} \frac{e^{z \log \log y}}{z^k} dz - \frac{\ell'(\rho)}{2\pi i} \rho \int_{|z|=\rho} \frac{e^{z \log \log y}}{z^{k+1}} dz \\ &= \ell'(\rho) \frac{(\log \log y)^{k-1}}{(k-1)!} - \ell'(\rho) \rho \frac{(\log \log y)^k}{k!}. \end{aligned} \tag{57}$$

With the choice of $\rho = \frac{k}{\log \log y}$, (57) becomes

$$I_2 = 0. \tag{58}$$

Equation (58) demonstrates why choosing $\rho = \frac{k}{\log \log y}$ is the best possible choice. Any other value for ρ in (58) would contribute an error term which would not be zero.

It remains to estimate the third integral I_3 . Let us set $z = \rho e^{2\pi i \theta}$ for $-\frac{1}{2} \leq \theta \leq \frac{1}{2}$, so that

$$z - \rho = 2i e^{\pi i \theta} \rho \frac{(e^{\pi i \theta} - e^{-\pi i \theta})}{2i} = 2i e^{\pi i \theta} \rho \sin \pi \theta \ll \rho \theta.$$

Thus

$$(z - \rho)^2 \ll \rho^2 \theta^2.$$

Next note that

$$\begin{aligned} |\log^z y| &= |e^{z \log \log y}| = e^{Re(z) \log \log y} = e^{\rho \cos 2\pi \theta \log \log y} \\ &\leq e^{\rho(1-\lambda \theta^2) \log \log y} = e^{k(1-\lambda \theta^2)} \end{aligned} \tag{59}$$

by again using the fact that $\rho = k / \log \log y$, and where $\lambda > 0$ is some small fixed constant. Applying the bound in (59) to the integral I_3 we see

$$\begin{aligned} I_3 &\ll \int_{-1/2}^{1/2} \frac{x \rho^2 \theta^2}{\log(y)} \frac{e^k e^{-k\lambda \theta^2} r |e^{2\pi i \theta}|}{r^{k+1}} d\theta \\ &\ll \frac{x}{\log y} \frac{e^k}{\rho^{k-2}} \int_{-1/2}^{1/2} \theta^2 e^{-k\lambda \theta^2} d\theta \ll \frac{x}{\log(y)} \frac{e^k}{r^{k-2}} \int_0^\infty \theta^2 e^{-k\lambda \theta^2} d\theta \end{aligned}$$

which after the substitution $\sqrt{k}\theta = u$ yields

$$\begin{aligned} & \frac{x}{\log(y)} \frac{e^k}{\rho^{k-2}} \int_0^\infty \frac{u^2 e^{-\lambda u^2}}{k\sqrt{k}} du \\ & \ll \frac{x}{\log y} \frac{(\log \log y)^{k-2}}{k^{k-2}} \frac{e^k}{k^{3/2}} \ll \frac{x}{\log y} \frac{(\log \log y)^{k-2}}{e^{-k} k^{k-1/2}} \\ & \ll \frac{x}{\log y} \frac{(\log \log y)^k}{k!} \frac{k}{(\log \log y)^2}, \end{aligned} \tag{60}$$

by using Stirling's formula once more. Combining (60) with the estimates for I_1 and I_2 gives

$$\begin{aligned} N_k(x, y) &= I_1 + I_2 + I_3 = \ell(\rho) \frac{x}{\log y} \frac{(\log \log y)^k}{k!} \left(1 + O\left(\frac{k}{(\log \log y)^2}\right) \right) \\ &= \ell\left(\frac{k}{\log \log y}\right) \frac{x}{\log y} \frac{(\log \log y)^k}{k!} \left(1 + O\left(\frac{k}{(\log \log y)^2}\right) \right). \end{aligned}$$

It is to be noted that for almost all integers we have $\omega_y(n) \sim \log \log y$, and for this situation

$$\frac{x}{\log y} \frac{(\log \log y)^k}{k!} \sim \frac{x}{\log y} \frac{(\log \log y)^{k-1}}{(k-1)!}.$$

On the other hand, if $k \sim \lambda \log \log y$ with $\lambda \neq 1$, then

$$\frac{x}{\log y} \frac{(\log \log y)^{k-1}}{(k-1)!} \sim \lambda \frac{x}{\log y} \frac{(\log \log y)^k}{k!},$$

that is, the terms differ by a factor of λ .

We will now study $N_k(x, y)$ for large y . Tenenbaum [14] has communicated to us that by suitably adapting the powerful techniques of [15] to $S_z(x, y)$, one may derive an effective estimate for $N_k(x, y)$ for certain ranges of k . These estimates follow from a suitable upper bound for $S_z(x, y)$ when $|z| = r$ but z is not close to r , and asymptotic estimates for $S_z(x, y)$ when z is close to r . Since the statement of his result involves $S_z(x, y)$ only when $z = r$ is real, we have $\rho = r$ as well. The form in which we will use his result is

Theorem 11 *Let $0 < \kappa < 1$ be a small parameter, $r = \frac{k}{\log \log y + c_1}$, where*

$$c_1 := \lim_{y \rightarrow \infty} \sum_{p < y} \frac{1}{p} - \log \log y.$$

Then for $\kappa \leq r \leq 1/\kappa$,

$$N_k(x, y) = S_r(x, y) \frac{(\log \log y + c_1)^k}{k!e^k} \left(1 + O\left(\frac{1}{\sqrt{\log \log y}}\right) \right)$$

uniformly for $1 \leq \alpha \leq (\log \log x)^2$.

Remark If one chooses $r = \frac{k}{\log \log y}$ in Theorem 11 then the result becomes

$$N_k(x, y) = S_r(x, y) \frac{(\log \log y)^k}{k!e^k} \left(1 + O\left(\frac{1}{\sqrt{\log \log y}}\right) \right) \tag{61}$$

which is more convenient.

Theorem 11 is interesting for several reasons. Firstly, it gives a relationship between the coefficients $N_k(x, y)$ of z^k in the sum $S_z(x, y)$ with the sum itself. Secondly, the size of sum $S_z(x, y)$ is utilized only when $z = r$ is real-valued and positive. Thankfully, the results of Sect. 4 apply in this situation and so we get the following theorem.

Theorem 12 Let $\kappa > 0$, $r = \frac{k}{\log \log y}$, and $\kappa \leq r \leq 1/\kappa$, then

$$N_k(x, y) = m_r(\alpha) \frac{x(\log \log y)^k}{k! \log y} \left(1 + O\left(\frac{1}{\sqrt{\log \log y}}\right) \right)$$

uniformly for $1 \leq \alpha \leq (\log \log x)^2$.

Proof By applying the estimate in Theorem 6 this with $\alpha \leq (\log \log x)^2$ in (61), we get

$$N_k(x, y) = m_r(\alpha) \frac{x(\log \log y)^k}{k!e^k \log^{1-r} y} \left(1 + O\left(\frac{1}{\sqrt{\log \log y}}\right) \right). \tag{62}$$

As $k = r \log \log y$, (62) can be rewritten in the form

$$N_k(x, y) = m_r(\alpha) \frac{x(\log \log y)^k}{k! \log y} \left(1 + O\left(\frac{1}{\sqrt{\log \log y}}\right) \right).$$

Remark We saw in Sect. 4 that when estimating $S_z(x, y)$ for large y , we were forced to restrict $Re(z) > 0$, so that a direct application of the Selberg contour integral method would not be possible without estimates for the case $Re(z) \leq 0$. However, with suitable bounds for $Re(z) \leq 0$ the method can still yield the correct asymptotic equality for $N_k(x, y)$. We note that the methods employed in Sect. 4 will provide us with the bound of

$$|S_z(x, y)| \ll \frac{x(\log \log x)^{K+1}}{\log y},$$

which is uniform provided $\alpha \ll \log \log x$, $Re(z) \leq 0$, and $|z| \leq R$ (and where K is the constant in Theorem 6). Note that this bound holds in the range $\alpha \ll \log \log x$ which overlaps with the range of α in Sect. 3. Therefore, we could have used this bound (which is milder than the results of Theorem 11). We made use of Theorem 11 to conclude the main result of Theorem 12 because of the sharpness of the error term in this theorem.

We note that when k is fixed, then $k/\log \log y$ is close to 0. But then $m_r(\alpha) \rightarrow m_0(\alpha) = w(\alpha)$, which means that Theorem 12 in this case corresponds asymptotically to Corollary 2. Also, if $y = x$ then Theorem 12 implies

$$\begin{aligned} N_k(x, y) &\sim \frac{S_r(x)}{e^k} \frac{(\log \log y)^k}{k!} = \frac{g(1, r)}{\Gamma(r)} \frac{x}{\log y} \frac{(\log \log y)^k}{k!} \\ &= \frac{g(1, r)r}{\Gamma(1+r)} \frac{x}{\log x} \frac{(\log \log x)^k}{k!} = \frac{g(1, r)}{\Gamma(1+r)} \frac{x}{\log x} \frac{(\log \log x)^{k-1}}{(k-1)!}, \end{aligned}$$

which corresponds to Selberg's estimate (3) for $N_k(x)$.

It is a matter of taste how one presents the estimate for $N_k(x, y)$. In Theorem 12, we wrote this in terms of ratios of elementary functions, and in the context of the results of the previous section, this is the more natural way to present this estimate. Tenenbaum has communicated Theorem 11 to us as a special case of his general result in [15] on the local distribution of $\omega(n; Q)$. However, as Q is a general set of primes, it is difficult to expect asymptotic estimates in terms of elementary continuous functions. Thus, Tenenbaum's result (Cor 2.4 of [15]) gives an estimate for the local distribution of $\omega(n; Q)$ in terms of $S_r(x, Q)$. If Q has regular behavior, such as when $Q = \{p|p < y\}$, then $S_r(x, Q) = S_r(x, y)$ can be estimated in terms of continuous functions (see [17]).

Acknowledgements The results presented here are contained in the PhD thesis of Todd Molnar submitted to the University of Florida in 2017 under the guidance of Krishnaswami Alladi. We are thankful to Professor Gerald Tenenbaum who generously shared his results in a letter. We are also appreciative of Professor Peter Elliott who went through the paper very carefully and suggested several edits which have improved the exposition.

References

- Alladi, K.: Asymptotic estimates of sums involving the Moebius function. *J. Number Theory* **14**, 86–98 (1982)
- Alladi, K.: Asymptotic estimates of sums involving the Moebius function. II. *Trans. Am. Math. Soc.* **272**, 87–105 (1982)
- Alladi, K.: The distribution of $\nu(n)$ in the sieve of Eratosthenes. *Q. J. Math. Oxf. (2)* **33**, 129–148 (1982)
- Alladi, K.: Multiplicative functions and Brun's Sieve. *Acta Arith.* **51**, 201–219 (1988)
- de Bruijn, N.G.: On the number of uncanceled elements in the sieve of Eratosthenes. *Indag. Math.* **12**, 247–256 (1950)
- Elliott, PDTA: *Probabilistic Number Theory*, vol. 1 and 2, Grundlehren vol. 239-240. Springer, Berlin (1979)
- Halász, G.: Remarks to my paper 'On the distribution of additive and the mean values of multiplicative arithmetic functions'. *Acta Math. Acad. Sci. Hung.* **23**, 425–432 (1972)

8. Hall, R.R., Tenenbaum, G.: Divisors, Cambridge Tracts in Mathematics, No. 90. Cambridge University Press, Cambridge (1988)
9. Hardy, G.H., Ramanujan, S.: The normal order of prime factors of a number n . Q. J. Math. (Oxf.) **48**, 76–92 (1917)
10. Landau, E: Handbuch der Lehre von der Verteilung der Primzahlen, Teubner, Leipzig, 3rd edn., vol. 2. Chelsea, New York (1974)
11. Sathe, LG: On a problem of Hardy on the distribution of integers with a given number of prime factors I–IV. J. Indian Math. Soc. (N.S.) **17**, 63–82 and 83–141, **18** 27–42 and 43–81
12. Selberg, A.: Note on a paper of L.G. Sathe. J. Indian Math. Soc. **18**, 83–87 (1954)
13. Tenenbaum, G.: Introduction to Analytic and Probabilistic Number Theory, Cambridge Studies in Advanced Mathematics 46. Cambridge University Press, Cambridge (1995)
14. Tenenbaum, G.: Private Communication (2016)
15. Tenenbaum, G.: Valeurs moyennes effectives de fonctions multiplicatives complexes. Ramanujan J. **44**, 641–701 (2017)
16. Tenenbaum, G: Generalized Mertens sums. In: G.E. Andrews, F. Garvan, (eds.) Analytic Number Theory, Modular Forms, and q -Hypergeometric Series—In Honor of Krishna Alladi's 60th Birthday, Springer Proc. in Math. and Stat., vol. 221, pp. 733–736 (2018)
17. Tenenbaum, G: On local laws for the number of small prime factors. Ramanujan J. (to appear)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.