**Michigan Math. J. 52 (2004)**

# Smooth Values of Shifted Primes in Arithmetic Progressions

WILLIAM D. BANKS, ASMA HARCHARRAS, & IGOR E. SHPARLINSKI

## 1. Introduction

Recall that an integer $n \geq 1$ is said to be *$y$-smooth* if it is not divisible by any prime $p$ exceeding $y$. Smooth numbers have played an important role in many number-theoretic and cryptographic investigations, and there is an extensive body of literature on the subject, originating with the work of Dickman [5] and de Bruijn [3]. For an interesting account of smooth numbers, we refer the reader to the survey article by Granville [1]; see also the references contained therein.

As usual, we denote by $\Psi(x, y)$ the counting function for smooth numbers:

$$\Psi(x, y) = \#\{n : 1 \leq n \leq x \text{ and } n \text{ is } y\text{-smooth}\}.$$

It is well known that the asymptotic relation

$$\Psi(x, y) \sim \rho(u)x$$

holds in a very wide range within the $xy$-plane, where $u = (\log x)/(\log y)$ and $\rho(u)$ is the Dickman–de Bruijn function defined by

$$\rho(u) = 1, \quad 0 \leq u \leq 1,$$

and

$$\rho(u) = 1 - \int_1^u \frac{\rho(v-1)}{v}\, dv, \quad u > 1.$$

For an account of the basic analytic properties of $\rho(u)$, we refer the reader to the book by Tenenbaum [19].

In this paper, we are interested in finding upper bounds for the number of primes $p \leq x$ that lie in a fixed arithmetic progression and such that $p - h$ is $y$-smooth, where $h \neq 0$ is a fixed integer.

To describe our results, let us introduce some notation that is used throughout the sequel. As usual, we denote by $\pi(x)$ the prime counting function:

$$\pi(x) = \#\{\text{prime } p \leq x\}.$$

Next, following [17], we define

Received June 4, 2003. Revision received January 20, 2004.
Work supported in part by NSF Grant no. DMS-0070628 for the first author and by ARC Grant no. DP0211459 for the third author.

$$\pi(x,y) = \#\{\text{prime } p \le x : p-1 \text{ is } y\text{-smooth}\}.$$

More generally, for any integer $h \neq 0$, let

$$\pi_h(x,y) = \#\{\text{prime } p \le x : p-h \text{ is } y\text{-smooth}\}.$$

Finally, for any integers $q, a$ with $q \ge 1$ and $\gcd(a,q) = 1$, we put

$$\pi_h(x,y;q,a) = \#\{\text{prime } p \le x : p-h \text{ is } y\text{-smooth and } p \equiv a \pmod q\}.$$

Since one might expect that the set $\{p-1 : \text{prime } p \le x\}$ contains roughly the same proportion of $y$-smooth integers as the set of all positive integers $n \le x$, it is reasonable to conjecture (following Erdős [6]; see also [16]) that the relation

$$\pi(x,y) \sim \rho(u)\pi(x)$$

holds for all $x$ and $y$ in a fairly wide range. More generally, for fixed $h \neq 0$ one might also conjecture (see e.g. [14]) that the relation

$$\pi_h(x,y) \sim \rho(u)\pi(x)$$

also holds in a wide range. Indeed, for all primes $p > |h|$, the quantity $p-h$ is relatively prime to $h$; thus it is reasonable to expect that the set $\{p-h : \text{prime } p \le x\}$ contains roughly the same proportion of $y$-smooth integers as the set of all positive integers $n \le x$ coprime to $h$, and for a fixed value of $h$ the latter proportion is easily seen to be $\rho(u)$ using a standard sieve to detect coprimality. Finally, arguing that the set of primes $p \le x$ such that $p-h$ is $y$-smooth is likely to be evenly distributed over all "admissible" arithmetic progressions modulo $q$ (i.e., those that represent infinitely many primes), it may be true that the relation

$$\pi_h(x,y;q,a) \sim \frac{\rho(u)}{\varphi(q)}\pi(x) \tag{1}$$

holds in a wide range. At the present time, however, all of these conjectures appear to be out of reach.

Over the years, substantial progress has been made on the problem of finding upper and lower bounds for $\pi(x,y)$; see [1; 2; 4; 6; 11; 14; 15; 16]. In particular, highly nontrivial upper bounds for $\pi(x,y)$ have been found by Fouvry and others in the case where $y > x^{1/2}$, while for smaller values of $y$, the bound

$$\pi(x,y) = O(u\rho(u)\pi(x))$$

has been obtained by Pomerance and Shparlinski [17] in the range

$$\exp\bigl(\sqrt{\log x \log\log x}\,\bigr) \le y \le x. \tag{2}$$

In the shorter range

$$\exp((\log x)^{2/3+\varepsilon}) \le y \le x, \tag{3}$$

the slightly stronger estimate

$$\pi(x,y) \ll \rho(u)\pi(x)$$

follows from Theorem 4 of Fouvry and Tenenbaum [8].

In this paper, we show how the methods of [17] combined with results of Granville [10] (see also [9]) and Fouvry and Tenenbaum [7] on smooth integers in arithmetic progressions can be used to prove that the estimate

$$\pi_h(x, y; q, a) = O\left(\frac{u\rho(u)c(h)}{\varphi(q)}\pi(x)\right) \tag{4}$$

holds uniformly (with respect to each of the involved parameters) for all $x$, $y$, and $q$ in a wide range, where

$$c(h) = \prod_{\substack{p \mid h \\ p \neq 2}} \frac{p-1}{p-2}.$$

Moreover, we obtain an explicit value (though somewhat modest) for the implied constant in (4). For a precise statement of this result, see Theorem 2. We remark that the conjecture (1) suggests that our upper bound (4) is probably not tight because it contains the extra factors $u$ and $c(h)$; it remains an interesting open question as to whether these factors can be removed from (4).

Our result immediately implies that rather sparse arithmetic progressions contain a positive proportion of primes for which $p-1$ has a large prime divisor; see the discussion in Section 5.

Our result also has an interesting cryptographic consequence. It is well known that primes $p$ for which $p-1$ is smooth are not suitable for most cryptographic applications derived from Diffie–Hellman or RSA schemes. The results of [17] show that, in fact, such primes are very rare. However, primes $p$ such that $p-1$ is not smooth are occasionally chosen to satisfy some additional conditions; for example, for the applications described in [20], what is needed is a good supply of nonsmooth primes lying in a certain arithmetic progression. The results of this paper imply, in a quantitative form, that almost all primes $p$ in any arithmetic progression with a modulus of moderate size are such that $p-1$ is not smooth.

Throughout the paper, the implied constants in the symbols $O$, $\ll$, and $\gg$ may occasionally depend, where obvious, on the small parameter $\varepsilon > 0$ but are absolute otherwise. We recall that the expressions $A \ll B$ and $B \gg A$ are each equivalent to the statement that $A = O(B)$. Throughout, the letters $p$ and $\ell$ always denote prime numbers, while $n$ and $q$ always denote positive integers.

ACKNOWLEDGMENTS. We wish to extend a special note of thanks to Glyn Harman, whose detailed comments on the manuscript helped to improve the exposition and the quality of our results. We also thank Roger Baker and Carl Pomerance for several enlightening discussions. The first two authors would like to thank Macquarie University for its hospitality during the preparation of this paper.

## 2. Preliminaries

Here we collect some preliminary estimates to be used in the sequel.

For any integer $n \geq 2$, let $P^+(n)$ denote the largest prime divisor of $n$, and put $P^+(1) = 1$. As in Section 1, we define

$$\Psi(x,y) = \#\{n \le x : P^+(n) \le y\}.$$

Throughout the sequel, the parameter $u$ is defined, as usual, to be the ratio $u = (\log x)/(\log y)$ whenever $x$ and $y$ are given.

The following result of Hildebrand [13] concerns the asymptotic nature of the function $\Psi(x,y)$; see also Corollary 9.3 in Chapter III.5 of [19].

LEMMA 1. *For every $\varepsilon > 0$, the estimate*

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

*holds uniformly provided that* $\exp((\log\log x)^{5/3+\varepsilon}) \le y \le x$.

For any integers $q$ and $a$ with $\gcd(a,q) = 1$, let

$$\Psi(x,y;q,a) = \#\{n \le x : P^+(n) \le y \text{ and } n \equiv a \pmod q\}$$

and put

$$\Psi_q(x,y) = \#\{n \le x : P^+(n) \le y \text{ and } \gcd(n,q) = 1\}.$$

We need the following result of Granville [10] about smooth numbers lying in a fixed arithmetic progression.

LEMMA 2. *For any $\varepsilon > 0$, the estimate*

$$\Psi(x,y;q,a) = \frac{1}{\varphi(q)}\Psi_q(x,y)\left(1 + O\left(\frac{\log q}{u^c \log y} + \frac{1}{\log y}\right)\right)$$

*holds uniformly, provided that* $\gcd(a,q) = 1$ *and* $q^{1+\varepsilon} \le y \le x$ *for some constant $c > 0$ that depends only on $\varepsilon$.*

Here, as usual, $\varphi$ denotes the Euler function.

We also need the following result of Fouvry and Tenenbaum [7] about smooth numbers relatively prime to a fixed modulus.

LEMMA 3. *For any $\varepsilon > 0$, the estimate*

$$\Psi_q(x,y) = \frac{\varphi(q)}{q}\Psi(x,y)\left(1 + O\left(\frac{\log\log(qy)\log\log x}{\log y}\right)\right)$$

*holds uniformly, provided that $x \ge x_0(\varepsilon)$,* $\exp((\log\log x)^{5/3+\varepsilon}) \le y \le x$, *and*

$$\log\log(q+2) \le \left(\frac{\log y}{\log(u+1)}\right)^{1-\varepsilon}.$$

We also need the following two lemmas concerning the Dickman–de Bruijn $\rho$-function.

LEMMA 4. *For any $u \ge 0$ and $0 \le \delta \le u$ with $\delta\log(u+1) \to 0$, the following estimate holds*:

$$\rho(u-\delta) = \rho(u)(1 + O(\delta\log(u+1))).$$

*Proof.* In Lemma 1 of [13], we find the estimate

$$-\frac{\rho'(u)}{\rho(u)} \le \log(u\log^2 u), \quad u \ge e^4.$$

In particular,

$$|\rho'(u)| \ll \rho(u)\log(u+1), \quad u \ge 0. \tag{5}$$

If $0 \le \delta \le u$ then, for some $v$ in the interval $[u-\delta, u]$,

$$\rho(u-\delta) - \rho(u) = \delta|\rho'(v)| \ll \delta\rho(v)\log(v+1) \le \delta\rho(u-\delta)\log(u+1);$$

that is,

$$\rho(u) = \rho(u-\delta)(1 + O(\delta\log(u+1))).$$

Taking into account that $\delta\log(u+1) \to 0$, we obtain the desired estimate. □

LEMMA 5. *For all $u \ge 1$,*

$$\int_{u-1}^{\infty} \rho(t)\,dt \le \left(u + 1 + \frac{1}{u}\right)\rho(u).$$

*Proof.* Using the well-known identity

$$\rho(u) = \frac{1}{u}\int_{u-1}^{u} \rho(t)\,dt, \quad u \ge 1$$

(see e.g. [13, Lemma 1]), it follows that $\rho(u) \le \rho(u-1)/u$; by induction, we obtain the estimate

$$\rho(u+j) \le \frac{\rho(u)}{(u+j)(u+j-1)\cdots(u+1)}, \quad j \ge 1.$$

Thus, for all $u \ge 1$, we have

$$\int_{u-1}^{\infty} \rho(t)\,dt = \sum_{j=0}^{\infty}(u+j)\int_{u+j-1}^{u+j}\rho(t)\,dt = \sum_{j=0}^{\infty}(u+j)\rho(u+j)$$

$$\le u\rho(u) + \rho(u) + \rho(u)\sum_{j=2}^{\infty}\frac{1}{(u+1)^{j-1}} = \left(u+1+\frac{1}{u}\right)\rho(u).$$

This completes the proof. □

Finally, we recall the following result from sieve theory; see [12, Thm. 3.12].

LEMMA 6. *Let $m$, $h$, $q$, and $b$ be integers such that*

$$mh \ne 0, \quad \gcd(m,h) = 1, \quad 2 \mid mh, \quad \gcd(q,b) = 1, \quad \text{and} \quad 1 \le q \le (\log Y)^A,$$

*where $A > 0$ is a fixed constant. Then, as $Y \to \infty$, the number of primes $\ell \le Y$ such that $\ell \equiv b \pmod q$ and $m\ell + h$ is prime is at most*

$$8\prod_{p>2}\left(1 - \frac{1}{(p-1)^2}\right)\prod_{2<p\mid qmh}\frac{p-1}{p-2}\,\frac{Y}{\varphi(q)\log^2 Y}\left(1 + O_A\left(\frac{\log\log Y}{\log Y}\right)\right)$$

*uniformly in the parameters $m$, $h$, $q$, and $b$.*

## 3. An Asymptotic Formula

Our principal tool is the following theorem, which is an extension of [17, Lemma 1] and which we believe to be of independent interest. While for the purposes of this paper we require only an upper bound for the sum considered below, we remark that Theorem 1 in fact provides an asymptotic formula in certain wide ranges of the involved parameters.

THEOREM 1. *Let $\varepsilon > 0$ be fixed. For any real numbers $x$ and $y$ satisfying*

$$\exp((\log\log x)^{5/3+\varepsilon}) \le y \le x,$$

*any integer $1 \le q \le y^{1-\varepsilon}$, and any integer $a$ with $\gcd(a,q) = 1$, we have*

$$\sum_{\substack{m \le x,\ P^+(m) \le y \\ m \equiv a \pmod q}} \frac{m}{\varphi(m)} = \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)} \frac{\rho(u)}{q} x(1 + O(\mathcal{E}_q(x,y))),$$

*where*

$$\mathcal{E}_q(x,y) = \frac{\log q}{\log y} + \frac{\log\log y \log\log x}{\log y} + \frac{y \log x \log\log x \log(u+1)}{x \log y},$$

*and $\zeta_q(s)$ is the partial zeta-function defined for $\Re(s) > 1$ by*

$$\zeta_q(s) = \prod_{p \nmid q} (1 - p^{-s})^{-1}.$$

*Proof.* For any integer $d$ with $\gcd(d,q) = 1$, denote by $d^*$ the unique integer such that $1 \le d^* \le q$ and $dd^* \equiv 1 \pmod q$. Then

$$\begin{aligned}
\sum_{\substack{m \le x,\ P^+(m) \le y \\ m \equiv a \pmod q}} \frac{m}{\varphi(m)} &= \sum_{\substack{m \le x,\ P^+(m) \le y \\ m \equiv a \pmod q}} \sum_{d \mid m} \frac{\mu(d)^2}{\varphi(d)} \\
&= \sum_{\substack{d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \sum_{\substack{m \le x/d,\ P^+(m) \le y \\ m \equiv ad^* \pmod q}} 1 \\
&= \sum_{\substack{d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \Psi(x/d, y; q, ad^*).
\end{aligned}$$

For the moment, suppose that $d \le x/y$. Since $q^{1+\varepsilon} \le y \le x/d$, Lemma 2 provides the uniform estimate

$$\Psi(x/d, y; q, ad^*) = \frac{\Psi_q(x/d, y)}{\varphi(q)} \left(1 + O\left(\frac{\log q}{u_d^c \log y} + \frac{1}{\log y}\right)\right)$$

for some constant $c > 0$ depending only on $\varepsilon$, where $u_d = (\log(x/d))/(\log y)$. If $q \ge 2$ then, since $u_d \ge 1$, we have

$$\Psi(x/d, y; q, ad^*) = \frac{\Psi_q(x/d, y)}{\varphi(q)}\left(1 + O\left(\frac{\log q}{\log y}\right)\right). \tag{6}$$

Clearly, this estimate holds also when $q = 1$.

Next, we want to apply Lemma 3 to estimate $\Psi_q(x/d, y)$. To do this, we need to check that the necessary conditions on $x/d$ and $y$ are met. First, observe that

$$\exp((\log\log(x/d))^{5/3+\varepsilon}) \le \exp((\log\log x)^{5/3+\varepsilon}) \le y \le x/d.$$

Also, $x/d \ge y \ge x_0(\varepsilon)$ if $x$ is sufficiently large. Finally, if $x$ and $y$ are large enough, then

$$u_d + 1 = \frac{\log(x/d)}{\log y} + 1 \le \frac{\log x}{(\log\log x)^{5/3+\varepsilon}} + 1 \le \log x,$$

$$\frac{\log y}{\log(u_d+1)} \ge \frac{\log y}{\log\log x} \ge \frac{\log y}{(\log y)^{1/(5/3+\varepsilon)}} \ge (\log y)^{2/5},$$

and therefore

$$\log\log(q+2) \le \log\log(y^{1-\varepsilon}+2) \le (\log y)^{2(1-\varepsilon)/5} \le \left(\frac{\log y}{\log(u_d+1)}\right)^{1-\varepsilon}.$$

Applying Lemma 3, we obtain the uniform estimate

$$\Psi_q(x/d, y) = \frac{\varphi(q)}{q}\Psi(x/d, y)\left(1 + O\left(\frac{\log\log(qy)\log\log(x/d)}{\log y}\right)\right).$$

Since

$$\log\log(qy) \le \log\log y^{2-\varepsilon} = O(\log\log y),$$

it follows that

$$\Psi_q(x/d, y) = \frac{\varphi(q)}{q}\Psi(x/d, y)\left(1 + O\left(\frac{\log\log y\log\log x}{\log y}\right)\right). \tag{7}$$

Combining the estimates (6) and (7), we now derive that

$$\Psi(x/d, y; q, ad^*) = \frac{\Psi(x/d, y)}{q}\left(1 + O\left(\frac{\log q}{\log y} + \frac{\log\log y\log\log x}{\log y}\right)\right) \tag{8}$$

provided that $d \le x/y$. For any $d > x/y$, we also have

$$\Psi(x/d, y; q, ad^*) = \sum_{\substack{n \le x/d \\ n \equiv ad^* \pmod q}} 1 = \frac{x}{dq} + O(1). \tag{9}$$

Now put $z = \min\{\log y, x/y\}$. Using (8) and (9), it follows that

$$\begin{aligned}\sum_{\substack{m \le x,\ P^+(m) \le y \\ m \equiv a \pmod q}} \frac{m}{\varphi(m)} &= \sum_{\substack{d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)}\Psi(x/d, y; q, ad^*) \\ &= (\Sigma_1 + \Sigma_2)\left(1 + O\left(\frac{\log q}{\log y} + \frac{\log\log y\log\log x}{\log y}\right)\right) \\ &\quad + O(\Sigma_3 + \Sigma_4),\end{aligned}$$

where

$$\Sigma_1 = \sum_{\substack{d \le z,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \frac{\Psi(x/d, y)}{q},$$

$$\Sigma_2 = \sum_{\substack{z < d \le x/y,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \frac{\Psi(x/d, y)}{q},$$

$$\Sigma_3 = \sum_{\substack{x/y < d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \frac{x}{dq},$$

$$\Sigma_4 = \sum_{\substack{x/y < d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)}.$$

First, let us estimate $\Sigma_1$. For all $d \le z$, by Lemma 1 we have

$$\begin{aligned}\Psi(x/d, y) &= \rho(u_d)\frac{x}{d}\left(1 + O\left(\frac{\log(u_d+1)}{\log y}\right)\right) \\ &= \rho\left(\frac{\log x - \log d}{\log y}\right)\frac{x}{d}\left(1 + O\left(\frac{\log\log x}{\log y}\right)\right).\end{aligned}$$

Also, by Lemma 4 we have

$$\rho\left(\frac{\log x - \log d}{\log y}\right) = \rho(u)\left(1 + O\left(\frac{\log\log y \log\log x}{\log y}\right)\right),$$

since if $\delta = (\log d)/(\log y)$ then

$$0 \le \delta \log(u+1) \le \frac{\log\log y \log\log x}{\log y},$$

and the last term tends to zero as $x \to \infty$ when $x$ and $y$ lie in the specified range. Thus, we derive that

$$\Sigma_1 = \frac{\rho(u)x}{q}\left(1 + O\left(\frac{\log\log y \log\log x}{\log y}\right)\right) \sum_{\substack{d \le z,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{d\varphi(d)}.$$

Finally, by the well-known inequality for the Euler function

$$\varphi(d) \gg \frac{d}{\log\log d}, \quad d \ge 3 \tag{10}$$

(see e.g. Theorem 5.1 from [18, Ch. 1]), we have

$$\begin{aligned}\sum_{\substack{d \le z,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{d\varphi(d)} &= \sum_{\substack{d \ge 1 \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{d\varphi(d)} + O\left(\sum_{d > \log y} \frac{\log\log d}{d^2}\right) \\ &= \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)} + O\left(\frac{\log\log\log y}{\log y}\right).\end{aligned}$$

Noting that

$$1 < \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)} < 2 \quad \forall q \geq 1, \tag{11}$$

it now follows that

$$\Sigma_1 = \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)} \frac{\rho(u)x}{q}\left(1 + O\left(\frac{\log\log y \log\log x}{\log y}\right)\right).$$

To estimate $\Sigma_2$, we may assume that $z = \log y < x/y$, since $\Sigma_2 = 0$ otherwise. Under this assumption, put

$$j_0 = \lfloor \log z \rfloor = \lfloor \log\log y \rfloor,$$
$$j_1 = \lfloor \log(x/y) \rfloor = \lfloor \log x - \log y \rfloor.$$

Then, using Lemma 1 again, we derive that

$$\begin{aligned}
\Sigma_2 &= \frac{1}{q} \sum_{\substack{\log y < d \leq x/y,\ P^+(d) \leq y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \Psi(x/d, y) \\
&\ll \frac{1}{q} \sum_{\log y < d \leq x/y} \frac{\mu(d)^2}{\varphi(d)} \rho\left(\frac{\log x - \log d}{\log y}\right) \frac{x}{d} \\
&\leq \frac{x}{q} \sum_{j_0 \leq j \leq j_1} \rho\left(u - \frac{j+1}{\log y}\right) \sum_{e^j < d \leq e^{j+1}} \frac{\mu(d)^2}{d\varphi(d)},
\end{aligned}$$

and therefore

$$\Sigma_2 \ll \frac{x}{q} \sum_{j_0 \leq j \leq j_1} \frac{\log j}{e^j} \rho\left(u - \frac{j+1}{\log y}\right). \tag{12}$$

Here we have used (10) and the fact that $\rho(u)$ is a decreasing function of $u$.

Next, we observe that the function

$$f(t) = \log(t-1) e^{-(t-1)} \rho\left(u - \frac{t}{\log y}\right)$$

satisfies the estimate

$$\frac{f(t+1)}{f(t)} = \frac{1}{e}\left(1 + O\left(\frac{\log\log x}{\log y}\right)\right) \tag{13}$$

for all $t$ in the range $\log y \leq t \leq \log x - 1$. Indeed, by Lemma 4 we have that, for all $t \leq \log x - 1$,

$$\rho\left(u - \frac{t+1}{\log y}\right) = \rho\left(u - \frac{t}{\log y}\right)\left(1 + O\left(\frac{\log\log x}{\log y}\right)\right),$$

since if $\delta = 1/\log y$ then

$$0 \leq \delta \log\left(u - \frac{t}{\log y} + 1\right) \leq \frac{\log\log x}{\log y} \to 0.$$

The estimate (13) follows immediately. This shows that $f(t)$ is a function of exponential decay provided that $x$ is sufficiently large; hence from (12) we now derive that

$$\Sigma_2 \ll \frac{x}{q} j_0 e^{-j_0} \rho\Big(u - \frac{j_0+1}{\log y}\Big).$$

Since $j_0 = \log\log y + O(1)$, we have

$$j_0 e^{-j_0} = O\bigg(\frac{\log\log y}{\log y}\bigg),$$

and another application of Lemma 4 then yields

$$\rho\Big(u - \frac{j_0+1}{\log y}\Big) = \rho(u)\bigg(1 + O\bigg(\frac{\log\log y \log\log x}{\log y}\bigg)\bigg).$$

Using (11), it follows that

$$\Sigma_2 \ll \frac{\rho(u)x}{q} \frac{\log\log y}{\log y}\bigg(1 + O\bigg(\frac{\log\log y \log\log x}{\log y}\bigg)\bigg) \ll \Sigma_1 \frac{\log\log y}{\log y}.$$

Finally, we turn to the estimates for $\Sigma_3$ and $\Sigma_4$. Put $j_2 = \lfloor \log x \rfloor$. Using (10), we derive that

$$\Sigma_3 = \frac{x}{q} \sum_{\substack{x/y < d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{d\varphi(d)} \ll \frac{x}{q} \sum_{\substack{x/y < d \le x \\ P^+(d) \le y}} \frac{\log\log d}{d^2}$$

$$\ll \frac{x}{q} \sum_{j_1 \le j \le j_2} \sum_{\substack{e^j < d \le e^{j+1} \\ P^+(d) \le y}} \frac{\log\log d}{d^2} \ll \frac{x}{q} \sum_{j_1 \le j \le j_2} \frac{\log j}{e^{2j}} \Psi(e^{j+1}, y).$$

By Lemma 1,

$$\Psi(e^{j+1}, y) = e^{j+1}\rho\Big(\frac{j+1}{\log y}\Big)\bigg(1 + O\bigg(\frac{\log\log x}{\log y}\bigg)\bigg),$$

and therefore

$$\Sigma_3 \ll \frac{x}{q} \sum_{j_1 \le j \le j_2} \frac{\log j}{e^j} \rho\Big(\frac{j+1}{\log y}\Big) \ll \frac{x}{q} \frac{\log j_1}{e^{j_1}} \rho\Big(\frac{j_1+1}{\log y}\Big) \ll \frac{y \log\log x}{q} \rho(u-1).$$

Since $\rho'(u) = -\rho(u-1)/u$ for all $u \ge 1$, the estimate (5) implies

$$\rho(u-1) \ll u \log(u+1)\rho(u), \quad u \ge 1.$$

Consequently,

$$\Sigma_3 \ll \frac{y \log\log x}{q} u \log(u+1)\rho(u) \ll \Sigma_1 \frac{y \log x \log\log x \log(u+1)}{x \log y}.$$

For $\Sigma_4$, we have the estimate

$$\Sigma_4 = \sum_{\substack{x/y < d \le x,\ P^+(d) \le y \\ \gcd(d,q)=1}} \frac{\mu(d)^2}{\varphi(d)} \ll \sum_{\substack{x/y < d \le x \\ P^+(d) \le y}} \frac{\log\log d}{d} \ll \frac{\log\log x}{x} \Psi(x, y).$$

By Lemma 1,

$$\Sigma_4 \ll \frac{\log\log x}{x}\rho(u)x\left(1+O\left(\frac{\log(u+1)}{\log y}\right)\right) \ll \Sigma_1 \frac{q\log\log x}{x}.$$

Combining our estimates for $\Sigma_j$ $(1 \le j \le 4)$ and using the fact that $q < y$, the result follows. □

COROLLARY 1. *Let $\varepsilon > 0$ be fixed. For any real numbers $x$ and $y$ satisfying*

$$\exp((\log\log x)^{5/3+\varepsilon}) \le y \le x,$$

*any integer $q$ with $\log q = o(\log y)$, and any integer $a$ with $\gcd(a,q) = 1$, we have*

$$\sum_{\substack{m\le x,\ P^+(m)\le y\\ m\equiv a \pmod q}} \frac{m}{\varphi(m)} \sim \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)}\frac{\rho(u)}{q}x.$$

*Proof.* Examining the structure of the error term in Theorem 1, it follows that $\mathcal{E}_q(x,y) = o(1)$ provided that

$$\exp((\log\log x)^{5/3+\varepsilon}) \le y \le \frac{x}{(\log\log x)^{1+\varepsilon}}$$

and $\log q = o(\log y)$, and so we obtain the required asymptotic formula. For larger values of $y$, one can argue directly as follows:

$$\sum_{\substack{m\le x,\ P^+(m)\le y\\ m\equiv a \pmod q}} \frac{m}{\varphi(m)} = \sum_{\substack{m\le x\\ m\equiv a \pmod q}} \frac{m}{\varphi(m)} - \sum_{\substack{m\le x,\ P^+(m)> y\\ m\equiv a \pmod q}} \frac{m}{\varphi(m)}.$$

For the first summation, it is easy to check that

$$\sum_{\substack{m\le x\\ m\equiv a \pmod q}} \frac{m}{\varphi(m)} = \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)}\frac{\rho(u)x}{q}\left(1+O\left(\frac{\log\log\log x}{\log x}\right)\right);$$

for the second summation, we have

$$\begin{aligned}\sum_{\substack{m\le x,\ P^+(m)> y\\ m\equiv a \pmod q}} \frac{m}{\varphi(m)} &\le \sum_{\substack{r\le \log\log x\\ \gcd(r,q)=1}} \sum_{\substack{y<p\le x\\ p\equiv ar^* \pmod q}} \frac{pr}{\varphi(pr)}\\ &\ll \frac{x}{\varphi(q)\log x}\sum_{r\le \log\log x}\frac{r}{\varphi(r)}\\ &\ll \frac{x\log\log(q+2)\log\log x\log\log\log\log x}{q\log x}\\ &= o\left(\frac{\rho(u)x}{q}\right).\end{aligned}$$

Here we have used the Brun–Titchmarsh theorem for the second inequality (see [12, Thm. 2.2]). This completes the proof. □

## 4. Main Results

As in Section 1, for all $n \ge 1$ we define

$$c(n) = \prod_{\substack{\ell \mid n \\ \ell \ne 2}} \frac{\ell - 1}{\ell - 2}.$$

Then we have the following trivial estimate:

$$c(n) \le \prod_{\substack{\ell \mid n \\ \ell \ne 2}} \frac{\ell}{\ell - 1} \cdot \prod_{\ell > 2} \frac{(\ell - 1)^2}{\ell(\ell - 2)} = \frac{n}{C_2 \gcd(n, 2)\varphi(n)}, \tag{14}$$

where $C_2$ is the "twin primes constant" given by

$$C_2 = \prod_{p > 2} \left(1 - \frac{1}{(p-1)^2}\right) = 0.6601618158\ldots.$$

THEOREM 2. *Let $\varepsilon > 0$ and $A > 0$ be fixed. Then, for any real numbers $x$ and $y$ satisfying*

$$\exp\bigl(\sqrt{\log x\,(\log\log x)^{1+\varepsilon}}\bigr) \le y \le x$$

*with $x \ge x_0(\varepsilon, A)$, the estimate*

$$\pi_h(x, y; q, a) \le \left(8\delta_h \frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)} + o(1)\right) \frac{(u + 1 + 1/u)\rho(u)c(qh)}{q} \pi(x)$$

*holds provided that $q \le (\log y)^A$, $\gcd(a, q) = 1$, and $h \ne 0$, where $\delta_h = 1$ if $h$ is even and $\delta_h = 1/2$ if $h$ is odd.*

*Proof.* In what follows, $\ell$ always denotes a prime number. Let

$$\pi_{h,\ell}(x; q, a) = \#\{p \le x : P^+(p - h) = \ell \text{ and } p \equiv a \pmod q\}.$$

Let $z = \exp((\log\log x)^2)$, and suppose that $z \le Y \le x$. Then, assuming that $x$ is sufficiently large,

$$\begin{aligned}
\pi_h(x, Y; q, a) - \pi_h(x, Y/e; q, a) &= \sum_{Y/e < \ell \le Y} \pi_{h,\ell}(x; q, a) \\
&= \sum_{Y/e < \ell \le Y} \sum_{\substack{m \le (x-h)/\ell,\ P^+(m) \le \ell \\ m\ell + h \text{ prime} \\ m\ell + h \equiv a \pmod q}} 1 \\
&= \sum_{\substack{1 \le b \le q \\ \gcd(b, q) = 1}} \sum_{\substack{Y/e < \ell \le Y \\ \ell \equiv b \pmod q}} \sum_{\substack{m \le (x-h)/\ell,\ P^+(m) \le \ell \\ m\ell + h \text{ prime} \\ mb + h \equiv a \pmod q}} 1 \\
&\le \sum_{\substack{1 \le b \le q \\ \gcd(b, q) = 1}} \sum_{\substack{m \le xe/Y,\ P^+(m) \le Y \\ mb + h \equiv a \pmod q}} \sum_{\substack{Y/e < \ell \le Y \\ \ell \equiv b \pmod q \\ m\ell + h \text{ prime}}} 1.
\end{aligned}$$

By Lemma 6, we have the estimate

$$\sum_{\substack{Y/e<\ell\le Y\\ \ell\equiv b \pmod q\\ m\ell+h \text{ prime}}} 1 \le (8C_2+o(1))\frac{c(qmh)Y}{\varphi(q)\log^2 Y}.$$

Note that the summation on the left-hand side is actually zero (for sufficiently large $x$) if $\gcd(m,h)>1$ or if both $m$ and $h$ are odd. Therefore, using the trivial inequality $c(qmh)\le c(qh)c(m)$ and the estimate (14), it follows that

$$\pi_h(x,Y;q,a)-\pi_h(x,Y/e;q,a) \le (8\delta_h+o(1))\frac{c(qh)Y}{\varphi(q)\log^2 Y}\sum_{\substack{1\le b\le q\\ \gcd(b,q)=1}}\sum_{\substack{m\le xe/Y,\ P^+(m)\le Y\\ mb+h\equiv a \pmod q}}\frac{m}{\varphi(m)}.$$

By Corollary 1, we have

$$\begin{aligned}\pi_h(x,Y;q,a)-\pi_h(x,Y/e;q,a) &\le (8\delta_h\eta_q+o(1))\frac{c(qh)Y}{\varphi(q)\log^2 Y}\sum_{\substack{1\le b\le q\\ \gcd(b,q)=1}}\rho\Big(\frac{\log x-\log Y+1}{\log Y}\Big)\frac{x}{qY}\\ &\le (8\delta_h\eta_q+o(1))\frac{c(qh)x}{q\log^2 Y}\rho\Big(\frac{\log x}{\log Y}-1\Big),\end{aligned} \tag{15}$$

where

$$\eta_q=\frac{\zeta_q(2)\zeta_q(3)}{\zeta_q(6)}.$$

Now let $j_0=\lfloor \log(y/z)\rfloor$. Using the estimate (15) yields

$$\begin{aligned}\pi_h(x,y;q,a) &\le \pi_h(x,z;q,a)+\sum_{j=0}^{j_0}(\pi_h(x,y/e^j;q,a)-\pi_h(x,y/e^{j+1};q,a))\\ &\le (1+o(1))\Psi(x,z;q,a)\\ &\quad+(8\delta_h\eta_q+o(1))\frac{c(qh)x}{q}\sum_{j=0}^{j_0}\frac{1}{(\log y-j)^2}\rho\Big(\frac{\log x}{\log y-j}-1\Big).\end{aligned}$$

Arguing as in [17, p. 341], we have the estimate

$$\sum_{j=0}^{j_0}\frac{1}{(\log y-j)^2}\rho\Big(\frac{\log x}{\log y-j}-1\Big)\le\frac{\rho(u-1)}{\log^2 y}+\frac{1}{\log x}\int_{u-1}^{\infty}\rho(t)\,dt. \tag{16}$$

If $y$ lies in the stated range then, as $x\to\infty$,

$$\frac{\rho(u-1)}{\log^2 y}=o\Big(\frac{u\rho(u)}{\log x}\Big);$$

the integral on the right-hand side of (16) can be estimated using Lemma 5. Finally, using (8) with $d=1$ together with Lemma 1, we have

$$\Psi(x,z;q,a) = \frac{\Psi(x,z)}{q}(1+o(1)) = \frac{x}{q}\rho\bigg(\frac{\log x}{(\log\log x)^2}\bigg)(1+o(1)).$$

The result follows. □

We remark that the bound (4) stated in the Introduction follows immediately from Theorem 2 using the trivial estimate $c(qh) \le c(q)c(h)$ together with (14).

## 5. Concluding Remarks

We note that the range of Theorem 2 is slightly shorter than that of [17] given by (2). However, one can easily extend Theorem 2 to the same range with $o(1)$ replaced by $O(1)$; in this case, we lose the explicit form of the statement.

The integral in Lemma 5 can be calculated precisely for certain values of $u$. For example,

$$\int_{u-1}^{\infty} \rho(t)\,dt = \begin{cases} e^{\gamma} + 1 - u & \text{if } 1 \le u \le 2, \\ e^{\gamma} + 3 - 2u + (u-1)\log(u-1) & \text{if } 2 \le u \le 3, \end{cases}$$

where $\gamma$ is the Euler–Mascheroni constant. Similar expressions can be obtained whenever $u$ is reasonably small and can be used in place of $(u + 1 + 1/u)\rho(u)$ in the statement of Theorem 2 to obtain better estimates. In this way, we find that

$$\frac{4\zeta(2)\zeta(3)}{C_2\zeta(6)} \int_{u-1}^{\infty} \rho(t)\,dt < 1, \quad u > u_0 = 3.3558619258\ldots,$$

and hence for any fixed $A > 0$ we have

$$\liminf_{x\to\infty} \min_{\substack{1\le q\le (\log x)^A \\ \gcd(a,q)=1}} \frac{\#\{p \le x : p \equiv a \pmod q,\ P^+(p-1) \ge x^{0.295}\}}{(\pi(x)/\varphi(q))} > \frac{1}{16},$$

where $P^+(n)$ is the largest prime divisor of $n$. We also remark that the pair $(0.295, 1/16)$ can be replaced with $(0.270, 1/2)$ or $(0.257, 2/3)$, for instance. For $q = 1$, the current record with the exponent 0.677 instead of 0.295 has been established in [2] (though one loses the positivity in the density of primes), and it seems likely that—by appropriately modifying the techniques of that paper—this stronger result might also be obtained for primes in arithmetic progressions. This has never been worked out explicitly, however, and doing so would necessarily entail many technical and tedious calculations; thus Theorem 2 provides a reasonably painless shortcut to (albeit weaker) results of the same general type. In fact, the range of $q$ is quite generous and coincides with the range for which unconditional results have been obtained on the asymptotic formula for primes in arithmetic progressions.

It is easy to see that any improvement of our principal tool, Lemma 6, will immediately lead to an improvement of Theorem 2 (both with respect to the constant 8 and the range of $q$). One can also try to prove an analogue of Theorem 1 for the sum

$$\sum_{\substack{m\le x,\ P^+(m)\le y\\ m\equiv a \pmod q}} c(m).$$

Using (14), one sees that this sum is bounded by

$$\frac{\zeta_q(2)\zeta_q(3)}{C_2\zeta_q(6)}\,\frac{\rho(u)}{q}x(1+O(\mathcal{E}_q(x,y))),$$

and any improvement in this estimate would lead to a corresponding improvement of Theorem 2.

In principle, it should be possible to improve the bound in Theorem 2 (by about a factor of $u$) using the approach of [8], but (as in the case of $q = 1$) only in a narrower range of $y$; compare (2) and (3). As we have already mentioned, any further progress toward closing the gap between (1) and Theorem 2 would be of great interest.

Finally, we remark that it should be possible via similar techniques to obtain analogues of our results for primes $p \le x$ in an arithmetic progression such that $kp + h$ is $y$-smooth.

## References


[1] W. R. Alford, A. Granville, and C. Pomerance, *There are infinitely many Carmichael numbers,* Ann. of Math. (2) 139 (1994), 703–722.

[2] R. C. Baker and G. Harman, *Shifted primes without large prime factors,* Acta Arith. 83 (1998), 331–361.

[3] N. de Bruijn, *On the number of positive integers $\le x$ and free of prime factors $> y$,* Nederl. Akad. Wetensch. Proc. Ser. A 54 (1951), 50–60.

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

[5] K. Dickman, *On the frequency of numbers containing prime factors of a certain relative magnitude,* Ark. Mat. Astr. Fys. 22 (1930), 1–14.

[6] P. Erdős, *On the normal number of prime factors of $p-1$ and some other related problems concerning Euler's $\phi$-function,* Quart. J. Math. Oxford Ser. (2) 6 (1935), 205–213.

[7] É. Fouvry and G. Tenenbaum, *Entiers sans grand facteur premier en progressions arithmetiques,* Proc. London Math. Soc. (3) 63 (1991), 449–494.

[8] ———, *Répartition statistique des entiers sans grand facteur premier dans les progressions arithmétiques,* Proc. London Math. Soc. (3) 72 (1996), 481–514.

[9] A. Granville, *Integers, without large prime factors, in arithmetic progressions. I,* Acta Math. 170 (1993), 255–273.

[10] ———, *Integers, without large prime factors, in arithmetic progressions. II,* Philos. Trans. Roy. Soc. London Ser. A 345 (1993), 349–362.

[11] ———, *Smooth numbers: Computational number theory and beyond,* Proc. MSRI Conf. Algorithmic Number Theory: Lattices, Number Fields, Curves, and Cryptography (Berkeley, 2000) (J. Buhler, P. Stevenhagen, eds.), Cambridge Univ. Press (to appear).

[12] H. Halberstam and H.-E. Richert, *Sieve methods,* London Math. Soc. Monogr., 4, Academic Press, London, 1974.

[13] A. Hildebrand, *On the number of positive integers $\leq x$ and free of prime factors $> y$*, J. Number Theory 22 (1986), 289–307.
[14] G. Martin, *An asymptotic formula for the number of smooth values of a polynomial,* J. Number Theory 93 (2002), 108–182.
[15] P. Moree, *A note on Artin's conjecture,* Simon Stevin 67 (1993), 255–257.
[16] C. Pomerance, *Popular values of Euler's function,* Mathematika 27 (1980), 84–89.
[17] C. Pomerance and I. Shparlinski, *Smooth orders and cryptographic applications,* Algorithmic number theory (Sydney, 2002) (C. Fieker, D. Kohel, eds.), Lecture Notes in Comput. Sci., 2369, pp. 338–348, Springer-Verlag, Berlin, 2002.
[18] K. Prachar, *Primzahlverteilung,* Springer-Verlag, Berlin, 1957.
[19] G. Tenenbaum, *Introduction to analytic and probabilistic number theory,* Cambridge Univ. Press, 1995.
[20] S. A. Vanstone and R. J. Zuccherato, *Short RSA keys and their generation,* J. Cryptology 8 (1995), 101–114.

W. D. Banks
Department of Mathematics
University of Missouri
Columbia, MO 65211

bbanks@math.missouri.edu

A. Harcharras
Department of Mathematics
University of Missouri
Columbia, MO 65211

harchars@math.missouri.edu

I. E. Shparlinski
Department of Computing
Macquarie University
Sydney, NSW 2109
Australia

igor@ics.mq.edu.au