Smoothing "Smooth" Numbers

John B. Friedlander *

Scarborough College, University of Toronto, Scarborough, Ontario M1C 1A4, Canada.

Andrew Granville †

University of Georgia, Athens GA 30602, USA

and

The Isaac Newton Institute for the Mathematical Sciences, 20 Clarkson Road, Cambridge CB2 OEH, UK.

Abstract: An integer is called *y-smooth* if all of its prime factors are $\leq y$. An important problem is to show that the y-smooth integers up to x are equi-distributed amongst short intervals. In particular, for many applications we would like to know that if y is an arbitrarily small, fixed power of x then all intervals of length $\sqrt{x}$, up to x, contain, asymptotically, the same number of y-smooth integers. We come close to this objective by proving that such y-smooth integers are so equi-distributed in intervals of length $\sqrt{x}y^{2+\varepsilon}$, for any fixed $\varepsilon > 0$.

* Partially supported by NSERC grant A5123

† An Alfred P. Sloan Research Fellow. Partially supported by the NSF.

Smoothing "Smooth" Numbers

John B. Friedlander and Andrew Granville

Introduction.

In this paper we will investigate the distribution, in short intervals, of integers without large prime factors . Good estimates, in such questions, have turned out to be surprisingly difficult. For instance, one might suppose that proving the existence of an integer free of prime factors $> x^{1/5}$, in an interval $(x, x+\sqrt{x}]$, would be easy, but this doesn't turn out to be the case.

In the absence of strong results for this problem, researchers have looked at certain variants. For example,

In what ranges can one prove that the expected asymptotic formula

$$\psi(x+z,y)-\psi(x,y) \sim \frac{z}{x}\ \psi(x,y) \tag{1}$$

holds ?

Unfortunately what has been proved has the (very severe) restriction that x/z is extremely small compared to x: Hildebrand (1986) proved (1) for

$$1 \le x/z \le y^{5/12} \tag{2}$$

with a wide range for y,

$$x \ge y \ge \exp\left((\log\log x)^{5/3+\varepsilon}\right). \tag{3}$$

Granville's method (1993), extending "Hildebrand's identity" (1986) from a sum over primes to a sum over integers with exactly k prime factors, can be used to improve the exponent $5/12$ in (2) to $1-\varepsilon$ for any fixed $\varepsilon > 0$, though not without considerable work.

Using the saddle point method, Hildebrand and Tenenbaum (1986) proved an asymptotic formula for $\psi(x+z,y)-\psi(x,y)$ in an even wider range for y, but never valid outside the range

$$1 \le x/z \le \exp\left((\log x)^{3/5}\right).$$

Even assuming the Riemann Hypothesis, one can only obtain

$$1 \le x/z \le \exp\left(\log x/(\log\log x)^3\right),$$

from their method.

However this still leaves us unable to prove (1) with $z = x^{3/4}$ for any $y \le x^{1/4}$; which is especially frustrating in view of the fact that there is a trivial argument (Friedlander and Lagarias, 1987, p.264) showing the existence of such integers under essentially that restriction, specifically

$$1 \le x/z \le (1-\varepsilon)y. \tag{4}$$

In the absence of wider ranges for (1), Friedlander and Lagarias (1987) proved the lower bound

$$\psi(x+z,y) - \psi(x,y) \gg \frac{z}{x}\ \psi(x,y), \tag{5}$$

for $y = x^{1/u}$, $1 \le x/z \le y^{1+c(1-1/u^3)}$ for some absolute constant $c > 0$. Again not as wide a range as one might hope but nonetheless sufficient to break the barrier of the (trivial) range (4). This result remains the strongest of its type for intervals $z \le x^{\frac{1}{2}}$. For longer intervals, Balog (1987) markedly improved the situation by showing that, for any $\varepsilon, \delta > 0$, $\psi(x + x^{\frac{1}{2}+\delta}, y) - \psi(x,y) > 0$ for any $y > x^{\varepsilon}$. In section 4 we briefly sketch modifications to his argument, leading to a proof of the following strengthened version (which also follows from our main result):

Corollary. *Fix $\delta > 0$, $\varepsilon > 0$. Then there exists $c = c(\varepsilon, \delta) > 0$ such that*

$$\psi(x+z,y) - \psi(x,y) \ge c\ \frac{z}{x}\ \psi(x,y)$$

provided that $x \ge z \ge x^{\frac{1}{2}+\delta}$, $x \ge y \ge x^{\varepsilon}$, and x is sufficiently large.

Friedlander and Lagarias (1987) also investigated in which ranges of y and z one could prove that the lower bound (5) holds for "almost all" x, rather than for all x. Recently Hildebrand and Tenenbaum (1991, Theorem 5.7) added new ingredients to the argument in (Friedlander and Lagarias, 1987) to show that, in essentially the same ranges, the lower bound (5) could be replaced by the asymptotic formula (1). Specifically they proved:

Theorem HT. *Under the assumptions*

$$x \geq y \geq \exp\left((\log x)^{\frac{5}{6}+\varepsilon}\right), \tag{6}$$

$$x \geq z \geq y \exp\left((\log x)^{\frac{1}{6}}\right), \tag{7}$$

we have

$$\psi(n+z,y) - \psi(n,y) = \frac{z}{n}\psi(n,y)\left\{1 + O\left(\frac{\log(u+1)}{\log y}\right)\right\},$$

$(x = y^u)$, *holds for all but* $O\left(x\exp(-(\log x)^{\frac{1}{6}-\varepsilon})\right)$ *positive integers* $n \leq x$.

(The earlier result of (Friedlander and Lagarias, 1987) had given the lower bound (5) subject to the same conditions (6), (7), with a slightly smaller exceptional set.)

Here we extend the range for which (1) is known to hold for all x.

Theorem. *Fix* $\varepsilon > 0$. *The estimate*

$$\psi(x+z,y) - \psi(x,y) = \frac{z}{x}\ \psi(x,y)\left\{1 + O\left(\frac{(\log\log y)^2}{\log y}\right)\right\} \tag{8}$$

holds uniformly for

$$x \geq y \geq \exp\left((\log x)^{\frac{5}{6}+\varepsilon}\right) \tag{6}$$

with

$$x \geq z \geq x^{\frac{1}{2}} y^2 \exp\left((\log x)^{\frac{1}{6}}\right). \tag{9}$$

The proof involves rewriting $\psi(x+z,y) - \psi(x,y)$ as a sum of many terms of this same form and then applying the Hildebrand-Tenenbaum "almost-all" result above. It is this procedure of reducing a result about an individual sequence to a result about an average over many sequences that is the "smoothing" referred to in the title. In fact this principle works more generally. A rather simpler version can be used to show that the Friedlander-Lagarias "almost all" version of Theorem HT implies the Corollary (Balog's result) and similar results for arithmetic progressions (and other sequences of integers) could be proven, were it not for the lack of the appropriate "almost all" results on which to build them.

In the proof of the Theorem we will use various crude upper bounds, which may doubtlessly be improved, and might lead to a better error term in (8).

The Corollary follows easily from our Theorem for, by taking $\theta = \min(\varepsilon, \delta/3)$,

$$\psi(x+z,y) - \psi(x,y) \geq \psi(x+z,x^\theta) - \psi(x,x^\theta) \sim \frac{z}{x}\ \psi(x,x^\theta) \gg \frac{z}{x}\ \psi(x,y),$$

since $\psi(x,x^\theta) \gg x$. More generally, if $y^* = \min(y, \{z/(x^{\frac{1}{2}} \exp(\log x)^{\frac{1}{6}})\})$ then

$$\psi(x+z,y) - \psi(x,y) \geq \psi(x+z,y^*) - \psi(x,y^*) \gg \frac{z}{x}\ \psi(x,y^*)$$

which provides a non-trivial lower bound in the range

$$z/x^{\frac{1}{2}},\ y \geq \exp\left((\log x)^{\frac{5}{6}+\varepsilon}\right).$$

It is somewhat annoying that the Theorem does not give (1) for

$$\psi\left(x + x^{\frac{1}{2}+\delta}, x^\varepsilon\right) - \psi(x,x^\varepsilon),$$

unless $\delta > 2\varepsilon$. We hope some modification or extension of our work will relax this restriction although a weaker condition of this type is already required in (Friedlander and Lagarias, 1987) and (Hildebrand and Tenenbaum, 1991).

A non-trivial lower bound in the range $z \geq x^{\frac{1}{2}+\delta}$, but including values of y as small as $\exp\left((\log x)^{\frac{2}{3}+\varepsilon}\right)$, has recently been given by Harman [Ha], by a sharpened and modified version of Balog's argument; this lower bound is made explicit, and improved by Lenstra, Pila, and Pomerance (1993). It is not out of the question that one might be able to modify the ideas in our proof to take more immediate advantage of the zeta-function techniques which underpin all of the above results and thereby avoid appealing to the "almost all" results. This would not only make the proof more direct but also increase the chances of lowering 5/6 to 2/3 in (6).

Our results do not give any indication of how to break the "$\sqrt{x}$ barrier": that is, proving that $\psi(x+\sqrt{x},y) - \psi(x,y) > 0$ when y is an arbitrarily small power of x. This is evidently the most challenging open problem in this area. Another, perhaps easier, annoyingly open problem is to give upper bounds for $\psi(x+z,y) - \psi(x,y)$ of the right size, in a wide range.

Notation: Throughout we let $P(m)$ and $Q(m)$ denote the largest and smallest prime factors of m, respectively. The letters p and q always denote primes. As usual ε and (except in the next section) c refer to positive absolute constants though they may change value as we proceed.

1. Very large x values.

As x gets very large compared to y, the integers free from prime factors $> y$ get increasingly scarce, so we might not expect (1) to hold. Indeed it makes more sense to compare them to the number of such integers between x and $2x$, rather than the number up to x, for instance by the equation

$$\psi(x+z,y)-\psi(x,y) \sim \frac{z}{x}\left\{\psi(2x,y)-\psi(x,y)\right\}. \tag{1)'$$

By (Hildebrand and Tenenbaum, 1986, Corollary 3) this is equivalent to (1) when $\log y/\log\log x \to \infty$. If we were to suppose that each of these $\psi(2x,y)-\psi(x,y)$ integers between x and $2x$ fall in the interval $(x, x+z]$ with 'probability' z/x then, by the Central Limit Theorem, we would expect (1)' to hold uniformly in the range

$$z\Big/\frac{x}{(\psi(2x,y)-\psi(x,y))}\log x \longrightarrow \infty.$$

Actually, since such integers become increasingly scarce as x gets very large compared to y, we might guess that

$$\psi(x+z,y)-\psi(x,y)=0 \text{ or } 1, \text{ only},$$

for even quite large values of z. This is equivalent to showing that if $c > a > x$ are integers, composed only of prime factors $\le y$, then $c-a > z$. We will deduce such a result from

Oesterlé and Masser's **"*abc*-conjecture":** *Fix $\varepsilon > 0$. If a, b, c are positive integers with $a+b=c$ then*

$$c \ll_\varepsilon \left(\prod_{p|abc} p\right)^{1+\varepsilon}.$$

Suppose $c > a$ are composed only of prime factors $\le y$. Then, applying the abc-conjecture (taking $b = c-a$), we obtain

$$c \ll_\varepsilon \left((c-a)\prod_{p\le y} p\right)^{1+\varepsilon}$$

so that

$$c-a \gg_\varepsilon c^{1-\varepsilon}\left(\prod_{p\le y} p\right)^{-1}.$$

Using the Prime Number Theorem we then deduce that, for any fixed $\varepsilon > 0$,

$$\psi\left(x + x^{1-\varepsilon}e^{-y}, y\right) - \psi(x,y) = 0 \text{ or } 1$$

for all $y < (1-\varepsilon)\log x$, if x is sufficiently large.

The best unconditional result of this type, due to Stewart and Yu (1991), is that

$$c \ll_\varepsilon \exp\left(\left(\prod_{p|abc} p\right)^{\frac{2}{3}+\varepsilon}\right).$$

By a similar argument we deduce that for any fixed $\varepsilon > 0$,

$$\psi\left(x + (\log x)^{\frac{3}{2}-\varepsilon}e^{-y}, y\right) - \psi(x,y) = 0 \text{ or } 1,$$

provided $y < (\frac{3}{2} - \varepsilon)\log\log x$, if x is sufficiently large.

Tijdeman (1973) showed that $\psi(x+z,y) - \psi(x,y) = 0$ or 1 for any $z < x/(\log x)^c$ where $c = \exp\left(k(y/\log y)^4\right)$. This gives a stronger result than that above for smaller y, for instance $y < (\log\log x)^{\frac{1}{4}}$.

2. Preparations.

We begin this section by recalling a few basic facts about $\psi(x,y)$. Define $\varrho(u) = 1$ for $0 < u \le 1$ and

$$u\varrho(u) = \int_{u-1}^{u} \varrho(t)dt \qquad \text{for } u > 1.$$

The key result in the area is that

$$\psi(x,y) \sim x\varrho(u) \qquad \text{for } x = y^u, \tag{2.1}$$

for any fixed $u > 0$ (Dickman), and, in fact, for a wide range of u values: the widest now known is (3), due to Hildebrand (1986).

The function ϱ is positive and decreasing for $u \ge 1$ and satisfies

$$\varrho(u) = u^{-u+o(u)} \tag{2.2}$$

(much sharper estimates are known). From its monotonicity and the integral delay equation we deduce that $u\varrho(u) < \varrho(u-1)$, and so

$$\varrho(u)/\varrho(u+K) > \prod_{k=1}^{[K]} (u+k) \geq u^{K-1}, \tag{2.3}$$

if $u, K \geq 1$.

We now give two lemmata that will be required for the proof of the Theorem.

Lemma 1. *If $x \geq z \geq y$ and $\Delta < \mu$, with $y^{-\frac{1}{2}} \leq \mu \leq \frac{1}{2}$, then*

$$\psi(x,y) - \psi(x, y-\Delta y) \ll \frac{\mu x}{\log y};$$
$$(\psi(x+z,y) - \psi(x,y)) - (\psi(x+z, y-\Delta y) - \psi(x, y-\Delta y)) \ll \frac{\mu z}{\log y}.$$

Proof: In the right hand side of the two equations above, we are counting integers $\leq x$ and between x and $x+z$, respectively, which have their largest prime factor, q, between $y - \Delta y$ and y. There can be no more than $x/q + 1 = O(x/q)$, respectively $z/q + 1 = O(z/q)$, such integers divisible by the prime q; so if q is between y and $y - \Delta y$ ($> y - \mu y > y/2$), then these quantities must be $\leq O(x/y)$ and $O(z/y)$ respectively. By the Brun-Titchmarsh Theorem the number of primes q in this interval is $O(\mu y/\log y)$, and the result follows.

Lemma 2. *Fix $\varepsilon > 0$. Suppose that $x^{1/2} < M < x^{3/4}$, with*

$$e^{(\log x)^{\frac{5}{6}+\varepsilon}} < y \leq x/2M \quad \text{and} \quad 2Mye^{(\log x)^{\frac{1}{6}}} \leq z \leq x.$$

Then, for any subset $\mathcal{M}_0$ of the integers in $[M, 2M]$, we have

$$\sum_{m \in \mathcal{M}_0} \left(\psi\left(\frac{x+z}{m}, y\right) - \psi\left(\frac{x}{m}, y\right) \right) = \frac{z}{x} \sum_{m \in \mathcal{M}_0} \psi\left(\frac{x}{m}, y\right) \left\{ 1 + O\left(\frac{\log(u+1)}{\log y}\right) \right\}$$
$$+ O\left(z e^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}} \right). \tag{2.4}$$

Proof: We shall show that

$$\psi\left(\frac{x+z}{m}, y\right) - \psi\left(\frac{x}{m}, y\right) = \frac{z}{x}\, \psi\left(\frac{x}{m}, y\right) \left\{ 1 + O\left(\frac{\log(u+1)}{\log y}\right) \right\} + O(1) \tag{2.5}$$

holds for all but $O\left(Me^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}}\right)$ integers m, $M \le m \le 2M$. Adding together (2.5) for each $m \in \mathcal{M}_0$ satisfying this equation, and adding to the error term in (2.4) an amount bounded by

$$\frac{z}{x}\,\psi\left(\frac{x}{m},y\right)+\psi\left(\frac{x+z}{m},y\right)-\psi\left(\frac{x}{m},y\right)\le\frac{z}{x}\cdot\frac{x}{m}+\frac{z}{m}+1\ll\frac{z}{M}$$

for each $m \in \mathcal{M}_0$ not satisfying (2.5), we obtain (2.4).

Now, let $\theta = \varrho(u)/\log y$ and $M_j = M(1+\theta)^j$ for $j = 0, 1, \ldots, J$, where $(1+\theta)^J \ge 2 > (1+\theta)^{J-1}$. Since $J \ll \theta^{-1} = (\log y)\varrho(u)^{-1} \ll e^{(\log x)^{\frac{1}{6}-\frac{2\varepsilon}{3}}}$, by (2.2), we need only prove that (2.5) holds for all but $O\left(Me^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}}\right)$ integers m in each interval $(M_j, M_{j+1}]$, and then adjust the constant c.

For $m \in (M_j, M_{j+1}]$, let $T = z/M_{j+1}$ so that

$$\psi\left(\frac{x+z}{m},y\right)-\psi\left(\frac{x}{m},y\right)=\psi\left(\left[\frac{x}{m}\right]+T,y\right)-\psi\left(\left[\frac{x}{m}\right],y\right)+O\left(\left|T-\frac{z}{m}\right|+1\right). \tag{2.6}$$

If $[x/m] = [x/m']$ then $|(x/m) - (x/m')| < 1$ and so $|m' - m| < mm'/x < M_{j+1}^2/x$. Therefore, each exceptional integer n in Theorem HT corresponds to $O(M^2/x)$ exceptional integers m here; and so, by Theorem HT (but with ε replaced by $\varepsilon/2$, with x by x/M_j, and with z by T), we have

$$\psi\left(\left[\frac{x}{m}\right]+T,y\right)-\psi\left(\left[\frac{x}{m}\right],y\right)=\frac{T}{x/m}\psi\left(\frac{x}{m},y\right)\left\{1+O\left(\frac{\log(u+1)}{\log y}\right)\right\} \tag{2.7}$$

for all but $O\left(Me^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}}\right)$ integers $m \in (M_j, M_{j+1}]$ (we remark that the 'M' in the bound for the size of the exceptional set here comes from $x/M \times M^2/x$).

Now, for $m \in (M_j, M_{j+1}]$, we have

$$\left|\frac{T}{x/m}-\frac{z}{x}\right|\psi\left(\frac{x}{m},y\right)=\left|T-\frac{z}{m}\right|\frac{\psi(x/m,y)}{x/m}\le\left|T-\frac{z}{m}\right|,$$

and

$$\left|T-\frac{z}{m}\right|\le\frac{z}{M_j}-\frac{z}{M_{j+1}}=\frac{\theta z}{M_{j+1}}\ll\frac{z}{x}\,\frac{\frac{x}{m}\varrho(u)}{\log y}\ll\frac{z}{x}\,\psi\left(\frac{x}{m},y\right)\frac{1}{\log y},$$

by (2.1) in the range given by Hildebrand. Thus (2.5) follows for those m satisfying (2.7), by inserting (2.7) into (2.6), and using the last two estimates.

3. Smoothing; the Proof of the Main Theorem.

Assume that x is sufficiently large, and let $\delta = u/4(u + 2\log\log y)$ throughout. Note that, since $4 < u < (\log x)^{1/6-\varepsilon} < (\log y)^{1/5}$, by hypothesis, we have

$$4 < \delta^{-1} < 4 + 2\log\log y, \quad \log(u/\delta) \ll \delta \log y, \quad \text{and} \quad y^\delta > e^{(\log x)^{5/6+\varepsilon}}.$$

Using (2.2) and the monotonicity of ρ, we have

$$y^{-\delta/2} < \varrho(u) < \varrho(2) = 1 - \log 2 < 1/2.$$

Together the above bounds imply the validity of ranges in estimates used below.

We approximate $\psi(x+z, y) - \psi(x, y)$ by Σ, the number of integers $n, P(n) \le y$, in $(x, x+z]$ which have a divisor $m > \sqrt{x}$ with $Q(m) > y^\delta$. Evidently any n counted by $\psi(x+z, y) - \psi(x, y)$ but not by Σ has a divisor $m > \sqrt{x}, P(m) \le y^\delta$, so that $m \le y^\delta\sqrt{x}$. For each such m there can only be $O(z/m)$ values of n, so that

$$0 \le \psi(x+z, y) - \psi(x, y) - \Sigma \ll z \sum_{\substack{\sqrt{x} < m \le y^\delta\sqrt{x} \\ P(m) \le y^\delta}} \frac{1}{m} \ll z\delta\varrho\left(\frac{u}{2\delta}\right)\log y \ll \frac{z\varrho(u)}{\log y},$$

using partial summation with (2.1) in the range (3), and the fact that ρ is non-increasing for the third inequality, then (2.3) (since $u/2\delta = 2u + 4\log\log y$) for the last inequality.

We shall estimate Σ by summing according to the size of the divisor m of n obtained by multiplying together the largest prime factors of n, in non-increasing order, stopping as soon as $m > \sqrt{x}$. By the definition of Σ we are guaranteed that such an m exists, is unique, and is composed only of prime factors between y^δ and y. From our construction we know that $m/Q(m) \le \sqrt{x}$. Let $\mathcal{M}$ denote the set of such integers, so that

$$\Sigma = \sum_{m \in \mathcal{M}} \left(\psi\left(\frac{x+z}{m}, Q(m)\right) - \psi\left(\frac{x}{m}, Q(m)\right)\right). \tag{3.1}$$

We split up this sum, according to the value of $Q(m)$, as follows: For each integer $j \ge 0$, we let $y_j = y^\delta(1+\varrho(u))^j$ and $y_J \ge y > y_{J-1}$. Let $\mathcal{M}_j$ denote the subset of $\mathcal{M}$ for which $Q(m) \in (y_{j-1}, y_j]$. For these m, the mth term in (3.1) equals

$$\psi\left(\frac{x+z}{m}, y_{j-1}\right) - \psi\left(\frac{x}{m}, y_{j-1}\right) + O\left(\frac{z}{m}\,\frac{\varrho(u)}{\delta\log y}\right), \tag{3.2}$$

by Lemma 1 with $\mu = \varrho(u)$, since

$$Q(m) - y_{j-1} \le y_j - y_{j-1} = \varrho(u) y_{j-1}.$$

Next we partition the sets $\mathcal{M}_j$ up into sets $\mathcal{M}_{ij}$, which consist of those m in $\mathcal{M}_j$ for which $2^{i-1}\sqrt{x} < m \le 2^i\sqrt{x}$. Now, since $y^\delta \le Q(m) \le y$, we have $u \le u' := \log x/\log Q(m) \le u/\delta$. We use this to apply Lemma 2 with $\mathcal{M}_0 = \mathcal{M}_{ij}$, given the ranges (6) and (9). Thus the contribution to (3.1) from the main term of (3.2) is

$$\sum_{j=1}^{J} \sum_{i=1}^{\left[\frac{\log y_j}{\log 2}\right]+1} \sum_{m\in\mathcal{M}_{ij}} \left\{ \psi\left(\frac{x+z}{m}, y_{j-1}\right) - \psi\left(\frac{x}{m}, y_{j-1}\right) \right\} =$$

$$= \sum_{j=1}^{J} \sum_{i=1}^{\left[\frac{\log y_j}{\log 2}\right]+1} \left\{ \frac{z}{x} \sum_{m\in\mathcal{M}_{ij}} \psi\left(\frac{x}{m}, y_{j-1}\right) \left\{1 + O\left(\frac{\log(u/\delta)}{\delta \log y}\right)\right\} + O\left(z e^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}}\right) \right\};$$

and, by Lemma 1, together with (2.1) in Hildebrand's range (3), this is

$$= \sum_{j=1}^{J} \sum_{i=1}^{[\frac{\log y_j}{\log 2}]+1} \left\{ \frac{z}{x} \sum_{m\in\mathcal{M}_{ij}} \psi\left(\frac{x}{m}, Q(m)\right) \left\{1 + O\left(\frac{\log(u/\delta)}{\delta \log y}\right)\right\} \right.$$

$$\left. + O\left(\frac{z}{x} \sum_{m\in\mathcal{M}_{ij}} \frac{\varrho(u) x}{m \log y_{j-1}}\right) + O\left(z e^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}}\right) \right\}.$$

Collecting the above estimates we find that

$$\psi(x+z, y) - \psi(x, y) = \frac{z}{x} \sum_{m\in\mathcal{M}} \psi\left(\frac{x}{m}, Q(m)\right) \left\{1 + O\left(\frac{\log(u/\delta)}{\delta \log y}\right)\right\} + O\left(\frac{z\varrho(u)}{\delta^2 \log y}\right), \tag{3.3}$$

since we have $\sum_{m\in\mathcal{M}} 1/m \ll \delta^{-1}$, and since, by (2.2),

$$\sum_{j=1}^{J} \log y_j \ll J \log y \ll \varrho(u)^{-1} \log^2 y \ll e^{-c(\log x)^{\frac{1}{6}-\frac{\varepsilon}{2}}} \varrho(u)/\log y.$$

Evidently $\psi(x, y)$ is equal to $\sum_{m\in\mathcal{M}} \psi(x/m, Q(m))$, plus the number of integers $n \le x, P(n) \le y$ without a divisor $m \in \mathcal{M}$. Provided that $n \ge x^{3/4}$, we are guaranteed that such n must have a divisor $r \ge x^{1/4}$, with $P(r) \le y^\delta$; and thus r can be assumed to be in the interval $[x^{1/4}, x^{1/4}y^\delta]$. Therefore the number of such n is

$$\ll x^{3/4} + \sum_{\substack{x^{1/4} < r \le x^{1/4} y^\delta \\ P(r) \le y^\delta}} \frac{x}{r} \ll x\delta\varrho(u/4\delta) \log y \ll x\varrho(u)/\log y,$$

by partial summation and (2.1), and then using (2.3) (since $u/4\delta = u + 2\log\log y$).

We deduce from this and (3.3) that

$$\psi(x+z,y) - \psi(x,y) = \frac{z}{x}\ \psi(x,y)\left\{1 + O\left(\frac{\log(u/\delta) + \delta^{-1}}{\delta \log y}\right)\right\},$$

and the result follows.

Remarks: Evidently this last estimate easily implies a smaller but more complicated error term in (8). One can do even better by giving a sharper upper bound for $\sum_{m\in\mathcal{M}} 1/m$. However these improvements do not lead to an error term $O(\log(u+1)/\log y)$ in (8), which would be the smallest possible.

This proof may be shortened by taking $\psi(2x,y)-\psi(x,y)$ instead of $\psi(x,y)$ throughout. For then the (corresponding) final result follows easily using (3.3) with $z = x$; and this also renders unnecessary the first statement of Lemma 1.

4. Modifications to Balog's proof.

In the introduction we showed how the Corollary follows easily from our Theorem. Now we sketch an alternate proof, modifying Balog's argument (1987):

Like Balog we consider a weighted sum $\sum_n d_n$ with $d_n = \sum_{m_1 m_2 | n} a_{m_1} a_{m_2}$ and, here,

$$a_m = \begin{cases} 1 & \text{if } M^{1-\eta} < m < M, \quad Q(m) > y^{\frac{1}{2}}, \quad P(m) \le y \\ 0 & \text{otherwise,} \end{cases}$$

where $\eta > 0$ and $M = x^{\frac{1}{2}-\frac{1}{4k}}$ for some positive integer k. Let $\mathcal{M}(s) = \sum\limits_m a_m m^{-s}$ and $L(s) = \sum\limits_{L_1<\ell\le L_2} \ell^{-s}$ where $L_1 = x^{\frac{1}{2k}}$, $L_2 = 2x^{\frac{1}{2k}+2\eta}$. We apply Perron's formula to the function $L(s)\mathcal{M}^2(s)$. Following Balog's argument (which uses van der Corput's theorem, the large values theorem, etc.) but replacing the error term $(\log x)^{-A}$ by $x^{-\delta'}$ in several places (in all cases this stronger error follows from the results quoted there) we deduce that

$$\sum_{x<n\le x+z} d_n = z\left(\sum_m a_m/m\right)^2 + O\left(z^{1-\eta'}\right), \tag{4.1}$$

holds for $z = x^{\frac{1}{2}+\frac{1}{8k}+\delta_0}$, $\delta_0 > 0$, for all $\eta < \eta_0$ with some $\eta_0 = \eta_0(k, \delta_0) > 0$ and some $\eta'(k, \delta_0, \eta) > 0$.

Now, for given $\varepsilon > 0$, $\delta > 0$ we can fix $k > \max\{1/4\delta, 1/\varepsilon\}, \delta_0 < \delta/2$, and $\eta < \varepsilon/4$. We then have $d_n \ll 1$, $\sum_m a_m/m \gg 1$ with the implied constants depending on δ and ε, and also $d_n \neq 0 \Rightarrow P(n) \leq y$. Therefore (4.1) yields the Corollary.

References

Balog, A. 1987 On the distribution of integers having no large prime factor. *Journeés Arith. Besançon 1985, Astérisque* **147/148**, 27–31.

Friedlander, J.B. & Lagarias, J.C. 1987 On the distribution in short intervals of integers having no large prime factor. *J. Number Theory* **25**, 249–273.

Granville, A. 1993 Integers, without large prime factors, in arithmetic progressions II. *This volume*

Harman, G. 1991 Short intervals containing numbers without large prime factors. *Math. Proc. Cambridge Phil. Soc.* **109**, 1–5.

Hildebrand, A. 1985 Integers free of large prime factors in short intervals. *Quart J. Math.* (Oxford) **36**, 57–69.

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

Hildebrand, A. & Tenenbaum, G. 1986 On integers free of large prime factors, *Trans. Amer. Math. Soc.* **296**, 265–290.

Hildebrand, A. & Tenenbaum, G. 1991 Integers without large prime factors. *preprint*.

Lenstra, H.W., Pila, J. & Pomerance, C. 1993 A hyperelliptic smoothness test. *This volume*

Stewart, C.L. & Yu, K. 1991 On the *abc* conjecture. *Math. Ann.* **29**, 225–230.

Tijdeman, R. 1973 On integers with many small prime factors. *Comp. Math.* **26**, 319–330.