

IMPROVED ERROR BOUNDS FOR THE FERMAT PRIMALITY TEST ON RANDOM INPUTS

JARED D. LICHTMAN AND CARL POMERANCE

ABSTRACT. We investigate the probability that a random odd composite number passes a random Fermat primality test, improving on earlier estimates in moderate ranges. For example, with random numbers to 2^{200} , our results improve on prior estimates by close to 3 orders of magnitude.

1. INTRODUCTION

Part of the basic landscape in elementary number theory is the Fermat congruence: If n is a prime and $1 \leq b \leq n - 1$, then

$$(1.1) \quad b^{n-1} \equiv 1 \pmod{n}.$$

It is attractive in its simplicity and ease of verification: using fast arithmetic sub-routines, (1.1) may be checked in $(\log n)^{2+o(1)}$ bit operations. Further, its converse (apparently) seldom lies. In practice, if one has a large random number n that satisfies (1.1) for a random choice for b , then almost certainly n is prime. To be sure, there are infinitely many composites (the Carmichael numbers) that satisfy (1.1) for all b coprime to n , see [1]. And in [2] it is shown that there are infinitely many Carmichael numbers n such that (1.1) holds for $(1 - o(1))n$ choices for b in $[1, n - 1]$. (Specifically, for each fixed k there are infinitely many Carmichael numbers n such that the probability a random b in $[1, n - 1]$ has $(b, n) > 1$ is less than $1/\log^k n$.) However, Carmichael numbers are rare, and if a number n is chosen at random, it is unlikely to be one.

We say n is a *probable prime to the base b* if (1.1) holds. A probable prime is either prime or composite, but the terminology certainly suggests that it is probably prime! Specifically, let $P(x)$ denote the probability that an integer n is composite given that

- (i) n is chosen at random with $1 < n \leq x$, n odd,
- (ii) b is chosen at random with $1 < b < n - 1$, and
- (iii) n is a probable prime to the base b .

It is known that if x is sufficiently large, then $P(x)$ is small. Indeed, Erdős and Pomerance [8, Theorem 2.2] proved that

$$(1.2) \quad P(x) \leq \exp(-(1 + o(1)) \log x \log \log x / \log \log x)$$

as $x \rightarrow \infty$. In particular, $\lim P(x) = 0$. Kim and Pomerance [10] replaced the asymptotic inequality of (1.2) with the weaker, but explicit, inequality

$$P(x) \leq (\log x)^{-197} \quad \text{for } x \geq 10^{10^5}$$

2000 *Mathematics Subject Classification.* Primary 11Y11; Secondary 11A51.
Key words and phrases. Fermat test, Miller–Rabin test, probable prime.

and gave numerical bounds on $P(x)$ for $10^{60} \leq x < 10^{10^5}$. In this paper we simplify the argument in [10] and obtain better upper bounds on $P(x)$ for $10^{60} \leq x \leq 10^{90}$, as seen in Figure 1. In particular, at the start of this range, our bound is over 700 times smaller.

FIGURE 1. New bounds on $P(x)$.

x	Bound on $P(x)$ in [10]	New bound on $P(x)$
10^{60}	$7.16E-2$	$1.002E-4$
10^{70}	$2.87E-3$	$1.538E-5$
10^{80}	$8.46E-5$	$2.503E-6$
10^{90}	$1.70E-6$	$4.304E-7$
10^{100}	$2.77E-8$	$7.798E-8$

The notation aEm means $a \times 10^m$.

With these methods, we also obtain new nontrivial bounds for $2^{40} \leq x < 10^{60}$, values of x smaller than the methods in [10] could handle. These results are included in Figure 2.

FIGURE 2. Upper bound on $P(2^k)$.

k	$P(2^k) \leq$	k	$P(2^k) \leq$	k	$P(2^k) \leq$
40	$4.306E-1$	140	$3.265E-3$	240	$1.017E-5$
50	$2.904E-1$	150	$1.799E-3$	250	$5.876E-6$
60	$1.848E-1$	160	$9.932E-4$	260	$3.412E-6$
70	$1.127E-1$	170	$5.505E-4$	270	$1.992E-6$
80	$6.728E-2$	180	$3.064E-4$	280	$1.169E-6$
90	$4.017E-2$	190	$1.714E-4$	290	$6.888E-7$
100	$2.388E-2$	200	$9.634E-5$	300	$4.080E-7$
110	$1.435E-2$	210	$5.447E-5$	310	$2.428E-7$
120	$8.612E-3$	220	$3.097E-5$	320	$1.451E-7$
130	$5.229E-3$	230	$1.770E-5$	330	$8.713E-8$

We compute the exact values of $P(x)$ for $x = 2^k$ with $3 \leq k \leq 36$. Additionally, we estimate $P(x)$ for $x = 2^k$ with $30 \leq k \leq 50$, using random sampling. Calibrating these estimates against the true values for $30 \leq k \leq 36$ suggest that the estimates are fairly close to the true values for $37 \leq k \leq 50$, and almost certainly within an order of magnitude from the truth.

A number n is called L -smooth if all of its prime factors are bounded above by L . The method of [10] first computes the contribution to $P(x)$ from numbers that are not L -smooth (for an appropriate choice for L), and then enters a complicated argument based on the asymptotic method of [8] for the contribution of the L -smooth numbers. In addition to small improvements made in the non- L -smooth case, our principal new idea is to use merely that there are few L -smooth numbers. For

this we use the upper bound method pioneered by Rankin in [15] for this problem, obtaining numerically explicit upper bounds on sums over L -smooth numbers, c.f. equation (3.5) and Remark 3.4. These upper bounds should prove useful in other contexts.

One possible way to gain an improvement is to replace the Fermat test with the strong probable prime test of Selfridge. Also known as the Miller–Rabin test, it is just as simple to perform and it returns fewer false positives. To describe this test, let $n > 1$ be an odd number. First one computes s, t with $n - 1 = 2^s t$ and t odd. Next, one chooses a number b , $1 \leq b \leq n - 1$. The number n passes the test (and is called a *strong probable prime to the base b*) if either

$$(1.3) \quad b^t \equiv 1 \pmod{n} \quad \text{or} \quad b^{2^i t} \equiv -1 \pmod{n} \quad \text{for some } i < s.$$

Every odd prime must pass this test. Moreover, Monier [12] and Rabin [14] have shown that if n is an odd composite, then the probability that it is a strong probable prime to a random base b in $[1, n - 1]$ is less than $\frac{1}{4}$.

Let $P_1(x)$ denote the same probability as $P(x)$, except that (iii) is replaced by (iii)' n is a strong probable prime to the base b .

Based on the Monier-Rabin theorem, one might assume that $P_1(x) \leq \frac{1}{4}$, but as noted in [4], this reasoning is flawed. However, in [5] and [9], something similar to $P_1(x) \leq \frac{1}{4}$ is shown. Namely, if $P'_1(2^k)$ is the analogous probability for odd k -bit integers, it is shown in [5], [9] that $P'_1(2^k) \leq \frac{1}{4}$ for all $k \geq 3$. We show below how our estimates can be used to numerically bound $P_1(x)$. In particular, the results here improve on the estimates of [9] up to 2^{300} .

NOTATION

We have (a, b) , $[a, b]$ as the greatest common divisor, least common multiple of the positive integers a, b , respectively. We use p and q to denote prime numbers, and p_i to denote the i th prime. For $n > 1$, we let $P^+(n)$ denote the largest prime factor of n . Let φ denote Euler's function, λ the Carmichael universal exponent function, ζ the Riemann zeta-function, $\text{Li}(x) = \int_2^x \frac{dt}{\log t}$, and $\vartheta(x) = \sum_{p \leq x} \log p$. In many instances, we take a sum over certain subsets of odd composite integers, in which cases we use \sum'_n to denote $\sum_{\substack{n \text{ odd,} \\ \text{composite}}}$.

2. PRELIMINARY LEMMAS

In this section, we prove some preliminary lemmas which are needed for the rest of the paper, and which may be of interest in their own right.

Lemma 2.1. *Given real numbers a, b and a nonnegative, decreasing function f on the interval $[a, b]$, we have that*

$$\int_{[a]}^b f(t) dt \leq \sum_{a \leq n \leq b} f(n) \leq f(a) + \int_a^b f(t) dt.$$

The proof is clear. Note that since $\sum_{a < n \leq b} f(n) \leq \sum_{a \leq n \leq b} f(n)$, we may apply the upper bound for the sum on the half open interval.

Lemma 2.2. *For $x \geq 2$, we have that*

$$\frac{x}{\zeta(2)} - \log x \leq \sum_{n \leq x} \frac{\varphi(n)}{n} \leq \frac{x}{\zeta(2)} + \log x.$$

Proof. The result holds for $2 \leq x < 18$, so assume $x \geq 18$. We have that

$$\begin{aligned}
 \sum_{n \leq x} \frac{\varphi(n)}{n} &= \sum_{n \leq x} \sum_{d|n} \frac{\mu(d)}{d} = \sum_{d \leq x} \frac{\mu(d)}{d} \sum_{n \leq x/d} 1 = \sum_{d \leq x} \frac{\mu(d)}{d} \left\lfloor \frac{x}{d} \right\rfloor \\
 (2.1) \quad &= x \sum_{d \leq x} \frac{\mu(d)}{d^2} - \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \\
 &= \frac{x}{\zeta(2)} - x \sum_{d > x} \frac{\mu(d)}{d^2} - \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\},
 \end{aligned}$$

where $\{ \}$ denotes the fractional part. By Lemma 2.1,

$$\begin{aligned}
 \sum_{d > x} \frac{\mu(d)}{d^2} &\leq \sum_{d > x} \frac{1}{d^2} \leq \frac{1}{x^2} + \int_x^\infty \frac{dt}{t^2} = \frac{1}{x^2} + \frac{1}{x}, \\
 (2.2) \quad \sum_{d > x} \frac{\mu(d)}{d^2} &\geq - \sum_{d > x} \frac{1}{d^2} \geq -\frac{1}{x^2} - \frac{1}{x}.
 \end{aligned}$$

Since $\sum_{1 < d \leq 18, \mu(d) \neq -1} \frac{1}{d} = \frac{367}{336} > 1.09$, we have

$$- \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \leq \sum_{\substack{d \leq x \\ \mu(d) = -1}} \frac{1}{d} \leq \sum_{1 < d \leq x} \frac{1}{d} - \sum_{\substack{1 < d \leq 18 \\ \mu(d) \neq -1}} \frac{1}{d} < \log x - 1.09.$$

Substituting this and (2.2) back into (2.1) gives

$$\sum_{n \leq x} \frac{\varphi(n)}{n} \leq \frac{x}{\zeta(2)} + \frac{1}{x} + 1 + \log x - 1.09 < \frac{x}{\zeta(2)} + \log x.$$

Similarly, direct computation shows that

$$\sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \leq \sum_{\substack{d \leq x \\ \mu(d) = 1}} \frac{1}{d} \leq \sum_{1 < d \leq x} \frac{1}{d} - \sum_{\substack{1 < d \leq 4 \\ \mu(d) \neq 1}} \frac{1}{d} < \log x - \frac{13}{12}.$$

and thus

$$\sum_{n \leq x} \frac{\varphi(n)}{n} \geq \frac{x}{\zeta(2)} - \frac{1}{x} - 1 - \log x + \frac{13}{12} > \frac{x}{\zeta(2)} - \log x.$$

□

Lemma 2.3. *For $x \geq 1$, we have that*

$$\frac{\log x}{\zeta(2)} + 1 - \frac{\log 2}{\zeta(2)} < \sum_{n \leq x} \frac{\varphi(n)}{n^2} \leq \frac{\log x}{\zeta(2)} + 1.$$

Proof. The inequalities are easily verified for $x < 40$, so assume $x \geq 40$. Partial summation gives

$$\sum_{n \leq x} \frac{\varphi(n)}{n^2} = \sum_{n \leq 39} \frac{\varphi(n)}{n^2} + \frac{1}{x} \sum_{n \leq x} \frac{\varphi(n)}{n} - \frac{1}{40} \sum_{n \leq 39} \frac{\varphi(n)}{n} + \int_{40}^x \frac{1}{t^2} \sum_{n \leq t} \frac{\varphi(n)}{n} dt.$$

Evaluating the two sums to 39 and using the upper and lower bounds in Lemma 2.2 for the sums to x and t , we obtain the stronger result,

$$\frac{\log x}{\zeta(2)} + 0.58 < \sum_{n \leq x} \frac{\varphi(n)}{n^2} < \frac{\log x}{\zeta(2)} + 0.82.$$

Note that the upper bound in the lemma is tight at $x = 1$ and the lower bound cannot be improved as $x \rightarrow 2^-$. \square

Lemma 2.4. *If $2 \leq y < x$ and $0 < c < 1$, then*

$$\sum_{y < p \leq x} p^{-c} < f(x, y),$$

where

$$f(x, y) := (1 + 2.3 \cdot 10^{-8}) \left(\text{Li}(x^{1-c}) - \text{Li}(y^{1-c}) + \frac{y^{1-c}}{\log y} \right) - \vartheta(y) \frac{y^{-c}}{\log y}.$$

Proof. We use the inequalities

$$(2.3) \quad \vartheta(x) < x \quad (0 < x \leq 10^{19}), \quad |x - \vartheta(x)| < \epsilon x \quad (x > 10^{19}),$$

where $\epsilon = 2.3 \times 10^{-8}$, see [6], [7], improving on recent work in [13] (also see [11, Proposition 2.1]). Let $f(t) = 1/(t^c \log t)$. By partial summation,

$$\sum_{y < p \leq x} p^{-c} = \sum_{y < p \leq x} f(p) \log p = \vartheta(x) f(x) - \vartheta(y) f(y) - \int_y^x \vartheta(t) f'(t) dt.$$

Note that (2.3) implies that $\vartheta(t) < (1 + \epsilon)t$ for all $t > 0$. Since $f'(t) < 0$ for $t \geq 2$, we have

$$\begin{aligned} \sum_{y < p \leq x} p^{-c} &< (1 + \epsilon) x f(x) - (1 + \epsilon) \int_y^x t f'(t) dt - \vartheta(y) f(y) \\ &= (1 + \epsilon) \left(\text{Li}(x^{1-c}) - \text{Li}(y^{1-c}) + y f(y) \right) - \vartheta(y) f(y), \end{aligned}$$

where we have integrated by parts and used that $\int f(t) dt = \text{Li}(t^{1-c})$. This completes the proof. \square

Lemma 2.5. *We have*

$$\begin{aligned} (i) \quad \sum_{n > y} \frac{1}{n^2} &< \frac{5}{3y} && \text{for } y > 0, \\ (ii) \quad \sum_{n \geq y} \frac{1}{n^3} &\leq \frac{4(\zeta(3) - 1)}{y^2} && \text{for } y > 1. \end{aligned}$$

Proof. The first claim is stated and proved in (4.7) in [10]. We proceed similarly for the second claim. When $1 < y \leq 2$, we have

$$\sum_{n \geq y} \frac{1}{n^3} = \sum_{n \geq 2} \frac{1}{n^3} = \zeta(3) - 1 = \frac{4(\zeta(3) - 1)}{4} \leq \frac{4(\zeta(3) - 1)}{y^2}.$$

When $2 < y \leq 3$, direct computation shows that

$$\sum_{n \geq y} \frac{1}{n^3} = \sum_{n \geq 3} \frac{1}{n^3} = \zeta(3) - 1 - \frac{1}{8} < \frac{4(\zeta(3) - 1)}{y^2}.$$

When $3 < y \leq 4$, direct computation shows that

$$\sum_{n \geq y} \frac{1}{n^3} = \sum_{n \geq 4} \frac{1}{n^3} = \zeta(3) - 1 - \frac{1}{8} - \frac{1}{27} < \frac{4(\zeta(3) - 1)}{y^2}.$$

When $y > 4$, by Lemma 2.1, direct computation shows that

$$\sum_{n \geq y} \frac{1}{n^3} \leq \frac{1}{y^3} + \int_y^\infty \frac{dt}{t^3} = \frac{1}{y^3} + \frac{1}{2y^2} < \frac{4(\zeta(3) - 1)}{y^2}.$$

□

3. THE BASIC METHOD

Let

$$\mathbf{F}(n) = \{b \in (\mathbb{Z}/n\mathbb{Z})^\times : b^{n-1} = 1\}$$

and let $F(n) = \#\mathbf{F}(n)$. If $n > 1$ is odd, then $\pm 1 \in \mathbf{F}(n)$. Thus, for these n , $F(n) - 2$ counts the number of integers b , $1 < b < n - 1$, with $b^{n-1} \equiv 1 \pmod{n}$. Also note that by Fermat's little theorem, $F(p) = p - 1$ for primes p . We thus have for $x \geq 5$,

$$(3.1) \quad P(x) = \frac{\sum'_{n \leq x} (F(n) - 2)}{\sum_{1 < n \leq x, n \text{ odd}} (F(n) - 2)} = \left(1 + \frac{\sum_{2 < p \leq x} (p - 3)}{\sum'_{n \leq x} (F(n) - 2)}\right)^{-1}.$$

Hence to obtain an upper bound for $P(x)$, we shall be interested in obtaining a lower bound for $\sum_{2 < p \leq x} (p - 3)$ and an upper bound for $\sum'_{n \leq x} (F(n) - 2)$. To this end, we shall prove two theorems.

Theorem 3.1. *For $x \geq 2657$, we have*

$$\sum_{2 < p \leq x} (p - 3) > \frac{x^2}{2 \log x - \frac{1}{2}}.$$

Theorem 3.2. *Suppose c , L_1 , and L are arbitrary real numbers with $0 < c < 1$, $1 < L_1 < L$. Then for any $x > L^2$, we have*

$$\sum'_{n \leq x} (F(n) - 2) < x^{c+1} \prod_{2 < p \leq L} (1 - p^{-c})^{-1} + x^2 B,$$

where

$$B = \frac{1}{4L_1} + \frac{\log L_1}{\zeta(2)} \left(\frac{1}{2(L-1)} + \frac{1}{x^{1/2}} \right) + \frac{.5}{L-1} + \frac{.8}{x^{1/2}} \\ + \frac{L_1}{(x^{1/2} - 1)^2} + \frac{(1 + \log L_1)}{2(x^{1/2} - 1)} + \frac{1}{(L-1)^2} \left(\frac{L_1}{\zeta(2)} + \log L_1 \right).$$

Before proving Theorems 3.1 and 3.2, we state the main result of the section, which follows from these theorems.

Theorem 3.3. *Suppose c , L_1 , and L are arbitrary positive real numbers satisfying $0 < c < 1$ and $1 < L_1 < L$. Then for any $x > \max\{L^2, 2657\}$, we have $P(x) \leq 1/(1 + z^{-1})$ where*

$$z = \left(B + x^{c-1} \prod_{2 < p \leq L} (1 - p^{-c})^{-1} \right) (2 \log x - \frac{1}{2}),$$

and B is defined as in Theorem 3.2.

In principle the prime sum is much larger than the composite sum, so the probability $P(x)$ may be approximately viewed as their quotient. We remark that the prime sum in Theorem 3.1 is asymptotically equal to $x^2/(2 \log x)$, so the result is close to best possible. Additionally, in the application of Theorem 3.2, L and c are used as parameters for smoothness and Rankin's upper bound, respectively.

We now prove Theorem 3.1 using (2.3) and the additional inequalities from [6], [7] that

$$(3.2) \quad \vartheta(x) > x - 2\sqrt{x} \quad (1423 \leq x \leq 10^{19}), \quad \pi(x) < (1 + \epsilon)\text{li}(x) \quad (x \geq 2),$$

where $\text{li}(x) = \int_0^x dt/\log t$ and $\epsilon = 2.3 \times 10^{-8}$.

Proof of Theorem 3.1. Let $A = 1500$. By partial summation,

$$(3.3) \quad \begin{aligned} \sum_{2 < p \leq x} (p - 3) &= 1 - 3\pi(x) + \sum_{p \leq x} p \\ &= 1 - 3\pi(x) + \sum_{2 < p \leq A} (p - 3) + \frac{x\vartheta(x)}{\log x} - \frac{A\vartheta(A)}{\log A} - \int_A^x \vartheta(t) \frac{\log t - 1}{\log^2 t} dt. \end{aligned}$$

By (3.2), we have $-3\pi(x) > -3(1 + \epsilon)\text{li}(x)$. Suppose that $A \leq x \leq 10^{19}$. By (2.3), (3.2), we have

$$\begin{aligned} \frac{x\vartheta(x)}{\log x} - \int_A^x \vartheta(t) \frac{\log t - 1}{\log^2 t} dt &> \frac{x^2 - 2x^{3/2}}{\log x} - \int_A^x \frac{t}{\log t} - \frac{t}{\log^2 t} dt \\ &= \text{li}(x^2) - \frac{2x^{3/2}}{\log x} - \text{li}(A^2) + \frac{A^2}{\log A}. \end{aligned}$$

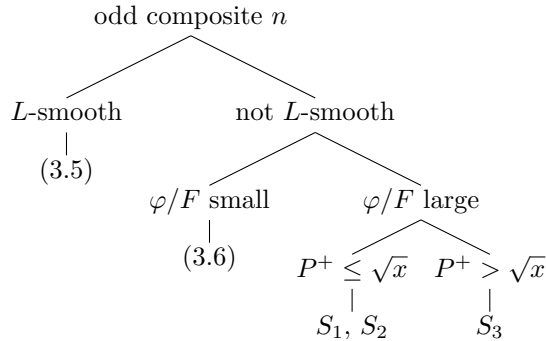
Using these estimates in (3.3), we have

$$\sum_{2 < p \leq x} (p - 3) > \text{li}(x^2) - \frac{2x^{3/2}}{\log x} - 3(1 + \epsilon)\text{li}(x) + 5875.$$

It is now routine to verify the theorem for $17000 \leq x \leq 10^{19}$. Similar calculations with (2.3), (3.2) establish the theorem for $x > 10^{19}$. A simple check then verifies the theorem in the stated range. \square

PROOF OF THEOREM 3.2

The bulk of the work is devoted to the proof of Theorem 3.2. The basic method is to divide the eligible n into five parts, depending on the largest prime factor $P^+(n)$ as well as the quotient $\varphi(n)/F(n)$, indicating how close n is to being a Carmichael number. We summarize this in the diagram below, which may help guide the reader through the proof.



For any $x > L^2$ with $L > L_1 > 1$, we have

$$(3.4) \quad \begin{aligned} \sum'_{n \leq x} (F(n) - 2) &= \sum'_{\substack{n \leq x \\ P^+(n) \leq L}} (F(n) - 2) + \sum'_{\substack{n \leq x \\ P^+(n) > L}} (F(n) - 2) \\ &\leq \sum'_{\substack{n \leq x \\ P^+(n) \leq L \\ n \text{ odd}}} n + \sum'_{\substack{n \leq x \\ P^+(n) > L}} F(n). \end{aligned}$$

For the first term in (3.4), we have for any $0 < c < 1$,

$$(3.5) \quad \sum_{\substack{n \leq x \\ P^+(n) \leq L \\ 2 \nmid n}} n \leq x^{1+c} \sum_{\substack{P^+(n) \leq L \\ 2 \nmid n}} \frac{1}{n^c} = x^{1+c} \prod_{2 < p \leq L} (1 - p^{-c})^{-1}.$$

Remark 3.4. By approximating the logarithm of the Euler product in (3.5) (with 2 included) using Lemma 2.4 and the method of [10], we can write a closed, numerically explicit upper bound on the distribution of L -smooth numbers: If $\frac{1}{2} < c < 1$ and $37 \leq L < x$, then

$$\sum_{\substack{n \leq x \\ P^+(n) \leq L}} 1 \leq x^c f_0 \exp(A + f(L, 36)),$$

where the notation $f(a, b)$ is defined in Lemma 2.4 and

$$f_0 := \prod_{p < 37} (1 - p^{-c})^{-1}, \quad A := \frac{1}{2c-1} \left(\frac{1}{2} + \frac{1}{3(37^c - 1)} \right) \left(36^{1-2c} - \frac{1}{2} \cdot 37^{1-2c} \right).$$

There has been a very recent improvement of this Rankin-type upper bound due to Granville and Soundarajan, see [11, Theorem 5.1], that is suitable for numerical estimates. It would be interesting to adapt that method to this paper.

Now we bound the second term in (3.4). Since $\mathbf{F}(n)$ is a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$, by Lagrange's Theorem we have $F(n) \mid \varphi(n)$, where φ is Euler's function. Then for each k , it makes sense to define $\mathbf{C}_k(x)$ as the set of odd, composite $n \leq x$ such that $F(n) = \varphi(n)/k$. Let $\mathbf{C}'_k(x)$ be the set of $n \in \mathbf{C}_k(x)$ for which $P^+(n) > L$, and let $C'_k(x) = \#\mathbf{C}'_k(x)$. Thus, we have

$$(3.6) \quad \begin{aligned} \sum'_{\substack{n \leq x \\ P^+(n) > L}} F(n) &= \sum_{k=1}^{\infty} \sum_{n \in \mathbf{C}'_k(x)} F(n) = \sum_{k=1}^{\infty} \sum_{n \in \mathbf{C}'_k(x)} \frac{\varphi(n)}{k} \\ &= \sum_{k \leq L_1} \frac{1}{k} \sum_{n \in \mathbf{C}'_k(x)} \varphi(n) + \sum_{k > L_1} \frac{1}{k} \sum_{n \in \mathbf{C}'_k(x)} \varphi(n) \\ &\leq x \sum_{k \leq L_1} \frac{C'_k(x)}{k} + \frac{1}{L_1} \sum_{\substack{1 < n \leq x \\ n \text{ odd}}} (n-2) \leq x \sum_{k \leq L_1} \frac{C'_k(x)}{k} + \frac{x^2}{4L_1}. \end{aligned}$$

It will thus be desirable to obtain an upper bound for $\sum_{k \leq L_1} \frac{C'_k(x)}{k}$. We remark that in the case $k > L_1$ we do not use $P^+(n) > L$; this observation will be useful in the next section.

Given a prime $p > L$, $d \mid p-1$, let

$$\mathbf{S}_{p,d}(x) = \{n : n \leq x \text{ odd, composite, } n \equiv p \pmod{p(p-1)/d}\}.$$

Let $S_{p,d}(x) = \#\mathbf{S}_{p,d}(x)$. Note that $S_{p,d} \leq \frac{xd}{p(p-1)}$ by the Chinese Remainder Theorem. We prove that

$$\bigcup_{k \leq L_1} \mathbf{C}'_k(x) \subset \bigcup_{\substack{d \leq L_1 \\ d \mid p-1 \\ L < p \leq x}} \mathbf{S}_{p,d}(x).$$

Take n in the left set. Then $p = P^+(n) > L$ and $k = \varphi(n)/F(n) \leq L_1$. By Lemma 2.4 in [10], we have $n \equiv 1 \pmod{\frac{p-1}{(k,p-1)}}$. Letting $d = (k, p-1)$, we have that $n \in \mathbf{S}_{p,d}$ (via the Chinese remainder theorem) and $d \leq k \leq L_1$, so n is in the right set.

Additionally, for a given p, d pair, $S_{p,d}$ counts integers $n = mp$ for which $m \equiv 1 \pmod{\frac{p-1}{d}}$. Write $m = 1 + u(\frac{p-1}{d})$ for some u . Letting $g = (u, d)$, we have that $m = 1 + (\frac{u}{g})(\frac{p-1}{d/g})$, so $n \in \mathbf{S}_{p,d/g}$, meaning that n will be counted multiple times if $g > 1$. Thus we require $(u, d) = 1$. In particular, if d is even, then u is odd. Since $m = 1 + u(\frac{p-1}{d})$ is odd, we have $u(\frac{p-1}{d})$ even. That is, if d is even then u is odd and $\frac{p-1}{d}$ is even, so $2d \mid p-1$. On the other hand, if d is odd, we of course have $2d \mid p-1$. Thus $2d \mid p-1$ always, and so

$$\begin{aligned} \sum_{k \leq L_1} \frac{C'_k(x)}{k} &\leq \sum_{d \leq L_1} \frac{1}{d} \sum_{\substack{L < p \leq x \\ 2d \mid p-1}} \sum_{\substack{u \leq \frac{xd}{p(p-1)} \\ (u,d)=1}} 1 \\ (3.7) \quad &= \sum_{d \leq L_1} \frac{1}{d} \sum_{\substack{L < p \leq x^{1/2} \\ 2d \mid p-1}} \sum_{\substack{u \leq \frac{xd}{p(p-1)} \\ (u,d)=1}} 1 + \sum_{d \leq L_1} \frac{1}{d} \sum_{\substack{x^{1/2} < p \leq x \\ 2d \mid p-1}} \sum_{\substack{u \leq \frac{xd}{p(p-1)} \\ (u,d)=1}} 1 \\ &< \sum_{d \leq L_1} \frac{\varphi(d)}{d} \sum_{\substack{L < p \leq x^{1/2} \\ 2d \mid p-1}} \left(\frac{x}{p(p-1)} + 1 \right) + \sum_{d \leq L_1} \sum_{\substack{x^{1/2} < p \leq x \\ 2d \mid p-1}} \frac{x}{n(n-1)} \\ &< S_1 + S_2 + S_3, \end{aligned}$$

where

$$\begin{aligned} (3.8) \quad S_1 &= x \sum_{d \leq L_1} \frac{\varphi(d)}{d} \sum_{\substack{L < n \leq x^{1/2} \\ 2d \mid n-1}} \frac{1}{(n-1)^2}, \quad S_2 = \sum_{d \leq L_1} \frac{\varphi(d)}{d} \sum_{\substack{1 < n \leq x^{1/2} \\ 2d \mid n-1}} 1, \\ S_3 &= \sum_{d \leq L_1} \sum_{\substack{x^{1/2} < n \leq x \\ 2d \mid n-1}} \frac{x}{(n-1)^2}. \end{aligned}$$

It is worth noting that in S_1, S_2, S_3 , we have dropped the condition that n be prime. An alternative bound using the condition of primality may be handled as an application of the Brun-Titchmarsh inequality. However, such a method is less effective for the small values of x considered here.

Consider S_1 in (3.8). For a given $d \leq L_1$, by Lemma 2.1 we have that

$$(3.9) \quad \sum_{\substack{n > L \\ 2d|n-1}} \frac{1}{(n-1)^2} = \sum_{2du+1 > L} \frac{1}{4d^2u^2} \leq \frac{1}{(L-1)^2} + \frac{1}{4d^2} \int_{(L-1)/2d}^{\infty} \frac{dt}{t^2} \\ = \frac{1}{(L-1)^2} + \frac{1}{2d(L-1)}.$$

Thus, by Lemma 2.2 and Lemma 2.3,

$$(3.10) \quad S_1 < x \sum_{d \leq L_1} \frac{\varphi(d)}{d} \left(\frac{1}{(L-1)^2} + \frac{1}{2d(L-1)} \right) \\ \leq \frac{x}{(L-1)^2} \left(\frac{L_1}{\zeta(2)} + \log L_1 \right) + \frac{x}{2(L-1)} \left(\frac{\log L_1}{\zeta(2)} + 1 \right).$$

By Lemma 2.3, S_2 in (3.8) is bounded by

$$(3.11) \quad S_2 \leq \sum_{d \leq L_1} \frac{\varphi(d)}{d} \frac{x^{1/2}}{2d} \leq x^{1/2} \left(\frac{\log L_1}{2\zeta(2)} + .8 \right).$$

We now consider S_3 in (3.8). For a fixed $d \leq L_1$, we have, as in (3.9),

$$\sum_{\substack{x^{1/2} < n \leq x \\ 2d|n-1}} \frac{1}{(n-1)^2} \leq \frac{1}{(x^{1/2}-1)^2} + \frac{1}{2d(x^{1/2}-1)}.$$

So,

$$(3.12) \quad S_3 \leq x \sum_{d \leq L_1} \frac{1}{(x^{1/2}-1)^2} + \frac{1}{2d(x^{1/2}-1)} \leq \frac{xL_1}{(x^{1/2}-1)^2} + \frac{x(1+\log L_1)}{2(x^{1/2}-1)}.$$

By (3.10), (3.11), and (3.12), we obtain from (3.7) that

$$(3.13) \quad \sum_{k \leq L_1} \frac{C'_k(x)}{k} < x \left(B - \frac{1}{4L_1} \right)$$

for B as in Theorem 3.2. Thus, using (3.13) in (3.6) gives the following result.

Theorem 3.5. *Suppose L_1 and L are arbitrary real numbers satisfying $1 < L_1 < L$. Then for any $x > L^2$, we have*

$$\sum'_{\substack{n \leq x \\ P^+(\bar{n}) > L}} (F(n) - 2) < x^2 B.$$

where B is as in Theorem 3.2.

Thus, (3.4), (3.5), and Theorem 3.5 give us Theorem 3.2.

4. A REFINEMENT OF THE BASIC METHOD

We refine the basic method as done analogously in [10], by considering the *two* largest prime factors of n . This refinement provides a modest improvement over Theorem 3.3 for x starting around 2^{140} .

Theorem 4.1. *Suppose c, L_1, L , and M are arbitrary real numbers satisfying $0 < c < 1$, $10 < L_1 < L$, $2L < M < L^2$. Then for any $x > L^2$, we have*

$$\sum'_{n \leq x} (F(n) - 2) < x^{c+1} (1 + f(L, M^{1/2})) \prod_{2 < p \leq M^{1/2}} (1 - p^{-c})^{-1} + x^2 (B + C),$$

where f is as in Lemma 2.4, B is as in Theorem 3.2, and

$$C = \frac{L^2}{2x} (1 + \log L_1) + \frac{2(1 + \log L_1)^2}{M \log M} + \frac{1}{12(M - 2L)} (1 + \log L) (4 + \log L_1)^4 \left(\frac{5}{12} + (\zeta(3) - 1)(1 + \log L) \right).$$

Proof. For each odd, composite $n \leq x$, letting P, Q be the two largest prime factors of n (i.e. $P = P^+(n), Q = P^+(n/P)$), we have three possible cases,

- (i) $P > L$ or $F(n) < \varphi(n)/L_1$,
- (ii) $P \leq L$ and $PQ \leq M$,
- (iii) $P \leq L, PQ > M$, and $F(n) \geq \varphi(n)/L_1$.

It is worth noting that cases (i) and (ii) are not in general mutually exclusive. We retain Theorem 3.5 and the remark following (3.6) to handle case (i). For case (ii), let $0 < c < 1$. When $P \leq M^{1/2}$, we have

$$\sum_{\substack{n \leq x, 2 \nmid n \\ P \leq M^{1/2}}} 1 \leq x^c \sum_{\substack{2 \nmid n \\ P^+(n) \leq M^{1/2}}} n^{-c}.$$

Similarly, when $P > M^{1/2}$ we have $Q \leq \frac{M}{P} < M^{1/2}$, so

$$\begin{aligned} \sum_{\substack{n \leq x, 2 \nmid n \\ M^{1/2} < P \leq L \\ Q \leq M^{1/2}}} 1 &\leq \sum_{M^{1/2} < p \leq L} \sum_{\substack{n \leq x/p \\ P^+(n) \leq M^{1/2} \\ 2 \nmid n}} 1 \leq \sum_{M^{1/2} < p \leq L} \sum_{\substack{P^+(n) \leq M^{1/2} \\ 2 \nmid n}} \left(\frac{x}{np} \right)^c \\ &= x^c \sum_{M^{1/2} < p \leq L} p^{-c} \sum_{\substack{2 \nmid n \\ P^+(n) \leq M^{1/2}}} n^{-c}. \end{aligned}$$

Using Lemma 2.4,

$$\begin{aligned} (4.1) \quad \sum_{\substack{n \leq x, 2 \nmid n \\ P \leq L \\ PQ \leq M}} 1 &\leq \sum_{\substack{n \leq x, 2 \nmid n \\ P \leq M^{1/2}}} 1 + \sum_{\substack{n \leq x, 2 \nmid n \\ M^{1/2} < P \leq L \\ Q \leq M^{1/2}}} 1 \\ &\leq x^c \sum_{\substack{2 \nmid n \\ P^+(n) \leq M^{1/2}}} n^{-c} + x^c \sum_{M^{1/2} < p \leq L} p^{-c} \sum_{\substack{2 \nmid n \\ P^+(n) \leq M^{1/2}}} n^{-c} \\ &= x^c \left(1 + \sum_{M^{1/2} < p \leq L} p^{-c} \right) \sum_{\substack{2 \nmid n \\ P^+(n) \leq M^{1/2}}} n^{-c} \leq x^c (1 + f(L, M^{1/2})) \sum_{\substack{2 \nmid n \\ P^+(n) \leq M^{1/2}}} n^{-c} \\ &= x^c (1 + f(L, M^{1/2})) \prod_{2 < p \leq M^{1/2}} (1 - p^{-c})^{-1}. \end{aligned}$$

We now have the following result.

Theorem 4.2. *If $0 < c < 1$, $1 < L < x$, and $L < M < L^2$, then*

$$\sum_{\substack{n \leq x, n \text{ odd} \\ P < L \\ PQ \leq M}} n \leq x^{c+1} (1 + f(L, M^{1/2})) \prod_{2 < p \leq M^{1/2}} (1 - p^{-c})^{-1},$$

where f is as in Lemma 2.4.

Consider n belonging to case (iii). For each k , let $\mathbf{B}_k(x)$ denote the set of such n with $\varphi(n)/F(n) = k$ and let $B_k(x) = \#\mathbf{B}_k(x)$. Thus,

$$(4.2) \quad \sum'_{n \text{ in case (iii)}} F(n) \leq x \sum_{k \leq L_1} \frac{B_k(x)}{k}.$$

By (2.11) in [8], we have $\lambda(n) \mid k(n-1)$ for all $n \in \mathbf{B}_k(x)$. Since $PQ \mid n$, we have $\lambda(PQ) \mid \lambda(n)$, so n satisfies the set of congruences

$$(4.3) \quad n \equiv 0 \pmod{PQ}, \quad k(n-1) \equiv 0 \pmod{\lambda(PQ)}.$$

Suppose first that $P = Q$. Then $\lambda(PQ) = P(P-1)$, so that (4.3) implies that $P \mid k$. For such a prime P , the number of $n \leq x$ with $P^2 \mid n$ is at most $x/P^2 < x/M$. Thus, the contribution for n in this case is at most

$$(4.4) \quad \frac{x}{M} \sum_{k \leq L_1} \frac{x}{k} \sum_{\substack{P \mid k \\ P > M^{1/2}}} 1 < \frac{x^2}{M} \left(\sum_{k \leq L_1} \frac{1}{k} \right) \frac{\log L_1}{\log M^{1/2}} < \frac{2x^2}{M \log M} (1 + \log L_1)^2.$$

Now consider the case $P > Q$. The latter congruence in (4.3) is equivalent to

$$n \equiv 1 \pmod{\left(\frac{\lambda(PQ)}{(k, \lambda(PQ))} \right)},$$

from which we also note

$$\left(PQ, \frac{\lambda(PQ)}{(k, \lambda(PQ))} \right) = 1.$$

Thus for arbitrary fixed primes $p > q$, the Chinese remainder theorem gives that the number of integers $n \leq x$ satisfying the system $n \equiv 0 \pmod{pq}$, $k(n-1) \equiv 0 \pmod{\lambda(pq)}$ as in (4.3) is at most

$$1 + \frac{x(k, \lambda(pq))}{pq\lambda(pq)}.$$

Summing over choices for p, q , we have the number of n in this case is at most

$$(4.5) \quad \sum_{\substack{q < p \leq L \\ pq > M}} \left(1 + \frac{x(k, \lambda(pq))}{pq\lambda(pq)} \right) \leq \frac{1}{2}L^2 + \frac{1}{2}x \sum_{\substack{p, q \leq L \\ pq > M \\ p \neq q}} \frac{(k, [p-1, q-1])}{pq[p-1, q-1]}.$$

This is (4.4) in [10] where “ L_2 ” there is our “ L ”. Following the argument in [10] from there, and letting $M' = M - 2L$ and with $u_1, u_2, u_3, u_4, \mu, \nu, \delta$ positive integer variables, we have that

$$(4.6) \quad \sum_{\substack{q, p \leq L \\ pq > M \\ p \neq q}} \frac{(k, [p-1, q-1])}{pq[p-1, q-1]} \leq \sum_{\substack{u_1 u_2 u_3 u_4 = k \\ (u_1, u_2) = 1}} \sum_{\substack{\mu \leq L/u_1 \\ \nu \leq L/u_2}} \sum_{u_1 u_2 u_3^2 \mu \nu \delta^2 > M'} \frac{1}{\mu^2 \nu^2 \delta^3 u_1 u_2 u_3^2}.$$

which is the initial inequality of (4.6) in [10] and with a typo corrected (the variable “ δ ” under the second summation there should be “ μ ”).

We now diverge from the argument in [10], and split up the sum on the right side of (4.6) into two cases, $\delta = 1$ and $\delta \geq 2$. When $\delta = 1$, by Lemma 2.5(i) we have

$$(4.7) \quad \sum_{\substack{u_1 u_2 u_3 u_4 = k \\ (u_1, u_2) = 1}} \sum_{\substack{\mu \leq L/u_1 \\ \nu \leq L/u_2}} \sum_{\mu \nu u_1 u_2 u_3^2 > M'} \frac{1}{\mu^2 \nu^2 u_1 u_2 u_3^2} < \frac{5}{3M'} \sum_{u_1 u_2 u_3 u_4 = k} \sum_{\nu \leq L/u_2} \frac{1}{\nu} \\ & \leq \frac{5}{3M'} (1 + \log L) \sum_{u_1 u_2 u_3 u_4 = k} 1,$$

When $\delta \geq 2$, let $D := \sqrt{M'/u_1 u_2 u_3^2 \mu \nu}$. By Lemma 2.5(ii) we have

$$(4.8) \quad \sum_{\substack{u_1 u_2 u_3 u_4 = k \\ (u_1, u_2) = 1}} \sum_{\substack{\mu \leq L/u_1 \\ \nu \leq L/u_2}} \sum_{\delta \geq \max\{2, D\}} \frac{1}{\mu^2 \nu^2 \delta^3 u_1 u_2 u_3^2} \leq \frac{4(\zeta(3) - 1)}{M'} \sum_{u_1 u_2 u_3 u_4 = k} \sum_{\substack{\mu \leq L/u_1 \\ \nu \leq L/u_2}} \frac{1}{\mu \nu} \\ & \leq \frac{4(\zeta(3) - 1)}{M'} (1 + \log L)^2 \sum_{u_1 u_2 u_3 u_4 = k} 1.$$

Substituting (4.7) and (4.8) back into (4.6) and then (4.5), we have

$$(4.9) \quad \sum_{k \leq L_1} \frac{1}{k} \sum_{\substack{q < p \leq L \\ pq > M}} \left(1 + \frac{x(k, \lambda(pq))}{pq \lambda(pq)} \right) < \frac{1}{2} L^2 (1 + \log L_1) \\ & + x(1 + \log L) \left(\frac{5}{6M'} + \frac{2(\zeta(3) - 1)}{M'} (1 + \log L) \right) \sum_{k \leq L_1} \frac{\tau_{(4)}(k)}{k},$$

where $\tau_{(i)}(k)$ is the number of ordered factorizations of k into i positive factors. In [10] (see (4.9)), an easy induction argument shows that

$$\sum_{k \leq y} \frac{\tau_{(i)}(k)}{k} \leq \frac{1}{i!} (i + \log y)^i$$

for any natural number i and any $y \geq 1$. Using this in (4.9) and then combining with (4.4) gives

$$x \sum_{k \leq L_1} \frac{B_k(x)}{k} \leq x^2 C,$$

where C is as in Theorem 4.1. Thus, from (4.2) we have the following result.

Theorem 4.3. *If $10 < L_1 < L < M/2$ and $x > L^2 > M$, then*

$$\sum'_{n \text{ in case (iii)}} F(n) \leq x^2 C,$$

where C is as in Theorem 4.1.

Combining Theorems 3.5, 4.2 and 4.3 yield Theorem 4.1. \square

Finally, Theorems 3.1 and 4.1 give the following result.

Theorem 4.4. *If $0 < c < 1$, $10 < L_1 < L$, $2L < M < L^2 < x$, and $x \geq 2657$, then $P(x) \leq 1/(1 + z^{-1})$ where*

$$z = \left(x^{c-1} (1 + f(L, M^{1/2})) \prod_{2 < p \leq M^{1/2}} (1 - p^{-c})^{-1} + B + C \right) (2 \log x - \frac{1}{2}),$$

f is as in Lemma 2.4, B is as in Theorem 3.2, and C is as in Theorem 4.1.

5. THE STRONG PROBABLE PRIME TEST

The next theorem extends the applicability of Theorems 3.3 and 4.4 to the probability, $P_1(x)$, that an odd composite $n \leq x$ passes the strong probable prime test to a random base. For an odd number n , let $S(n)$ denote the number of integers $1 \leq b \leq n - 1$ such that n is a strong probable prime to the base b , cf. (1.3). Thus,

$$P_1(x) = \frac{\sum'_{n \leq x} (S(n) - 2)}{\sum'_{n \leq x} (S(n) - 2) + \sum_{2 < p \leq x} (p - 3)}.$$

The following theorem together with Theorems 3.1, 3.2, and 4.1 allows for a numerical estimation of $P_1(x)$ for various values of x .

Theorem 5.1. *For $x \geq 1$, we have that*

$$\sum'_{n \leq x} (S(n) - 2) \leq \frac{1}{2} \sum'_{n \leq x} (F(n) - 2).$$

Proof. By (2.1) in [9], we have that $S(n) \leq 2^{1-\omega(n)}F(n)$, where $\omega(n)$ denotes the number of distinct prime factors of n . So, if n is odd and divisible by at least 2 different primes, we have $S(n) \leq \frac{1}{2}F(n)$. Further, if $n = p^a$ is an odd prime power then $S(p^a) = F(p^a) = p - 1$. Therefore we have

$$\begin{aligned} \sum'_{n \leq x} (S(n) - 2) &\leq \sum'_{n \leq x} \left(\frac{1}{2}F(n) - 2 \right) + \frac{1}{2} \sum_{\substack{2 < p^a \leq x \\ a \geq 2}} (p - 1) \\ &= \frac{1}{2} \sum'_{n \leq x} (F(n) - 2) - \sum'_{n \leq x} 1 + \frac{1}{2} \sum_{\substack{2 < p \leq x^{1/a} \\ a \geq 2}} (p - 1), \end{aligned}$$

so to prove the theorem it is enough to show that

$$(5.1) \quad \sum'_{n \leq x} 1 \geq \frac{1}{2} \sum_{\substack{2 < p \leq x^{1/a} \\ a \geq 2}} (p - 1).$$

Since 3 times an odd integer > 1 is an odd composite number, we have

$$\sum'_{n \leq x} 1 \geq \sum_{\substack{1 < m \leq x/3 \\ m \text{ odd}}} 1 = \left\lfloor \frac{x}{6} - \frac{1}{2} \right\rfloor > \frac{1}{6}x - \frac{3}{2}.$$

Also, since the primes larger than 2 are odd, for a given value of a we have

$$\frac{1}{2} \sum_{2 < p \leq x^{1/a}} (p - 1) \leq \sum_{j \leq \frac{1}{2}(x^{1/a} - 1)} j \leq \frac{1}{2} \left(\frac{1}{2}x^{1/a} - 1 \right) \left(\frac{1}{2}x^{1/a} + 1 \right) < \frac{1}{8}x^{2/a}.$$

Adding these inequalities for $a = 2, 3, \dots, \lfloor \log x / \log 3 \rfloor$, we see that (5.1) will follow if we show that

$$\frac{1}{6}x - \frac{3}{2} > \frac{1}{8}x + \frac{1}{8}x^{2/3} + \frac{1}{8}x^{1/2}(\log x / \log 3 - 3).$$

This inequality holds for $x \geq 254$. For $9 \leq x < 254$, (5.1) can be verified directly. Indeed, the prime sum in (5.1) increases only at the 8 powers of odd primes to 254 and it is enough to compute the two sums at those points. For $x < 9$,

$$\sum'_{n \leq x} (F(n) - 2) = \sum'_{n \leq x} (S(n) - 2) = 0,$$

so the theorem holds here as well. This completes the proof. \square

We remark that the same result holds for the Euler probable prime test (also known as the Solovay–Strassen test). This involves verifying that the odd number n satisfies $a^{(n-1)/2} \equiv \left(\frac{a}{n}\right) \pmod{n}$, where $\left(\frac{a}{n}\right)$ is the Jacobi symbol. Indeed, from Monier’s formula, see [8, (5.4)], we have that the number of bases $a \pmod{n}$ for which the Euler congruence holds is also $\leq 2^{1-\omega(n)}F(n)$. Like the strong test (as discussed in the introduction), an advantage with the Euler probable prime test is that more liars may be weeded out by repeating the test.

6. NUMERICAL RESULTS

We apply Theorems 3.3 and 4.4 to obtain numerical bounds on $P(x)$ for various values of x . In Figure 3, bounds on $P(2^k)$ are computed via Theorem 3.3 for $40 \leq k \leq 130$ and Theorem 4.4 for $140 \leq k \leq 330$, at which point the methods of this paper lose their edge over those in [10]. To select values for parameters L, L_1, M, c , we started with an initial guess based on [10], and then optimized each parameter in turn (holding the others fixed). The reported values were determined by repeated this process five times.

Note that the upper bounds in Theorems 3.3, 4.4 are decreasing functions in x , so one can use the Figure 3 data to compute upper bounds for values of x between consecutive entries.

We also compute the exact values of $P(x)$ for $x = 2^k$ when $k \leq 36$. By definition,

$$P(x) = \frac{S_c(x)}{S_c(x) + S_p(x)}$$

for

$$S_p(x) = \sum_{2 < p \leq x} (p - 3), \quad S_c(x) = \sum'_{n \leq x} (F(n) - 2).$$

For ease, we have split up the computation into dyadic intervals $(2^{k-1}, 2^k)$. Letting

$$S_p(x/2, x) = \sum_{x/2 < p \leq x} (p - 3), \quad S_c(x/2, x) = \sum'_{x/2 < n \leq x} (F(n) - 2),$$

we have that

$$(6.1) \quad P(2^k) = \frac{\sum_{j=3}^k S_c(2^{j-1}, 2^j)}{\sum_{j=3}^k S_p(2^{j-1}, 2^j) + S_c(2^{j-1}, 2^j)}.$$

FIGURE 3. Upper bound on $P(2^k)$.

k	L	L_1	$M^{1/2}$	c	$P(2^k) \leq$
40	307 ⁻	135		0.5440	4.306E-1
50	727 ⁻	318		0.5831	2.904E-1
60	1.860E+3	831		0.6235	1.848E-1
70	4.000E+3	1.75E+3		0.6491	1.127E-1
80	8.500E+3	3.72E+3		0.6704	6.728E-2
90	1.804E+4	7.55E+3		0.6906	4.017E-2
100	3.505E+4	1.54E+4		0.7052	2.388E-2
110	7.351E+4	3.27E+4		0.7217	1.435E-2
120	1.354E+5	5.95E+4		0.7321	8.612E-3
130	2.507E+5	1.10E+5		0.7423	5.229E-3
140	9.90E+5	1.57E+5	2.379E+5	0.7444	3.265E-3
150	2.20E+6	3.19E+5	3.739E+5	0.7504	1.799E-3
160	4.88E+6	6.21E+5	5.689E+5	0.7554	9.932E-4
170	1.05E+7	1.21E+6	8.669E+5	0.7602	5.505E-4
180	2.21E+7	2.30E+6	1.315E+6	0.7648	3.064E-4
190	4.55E+7	4.55E+6	1.990E+6	0.7692	1.714E-4
200	9.23E+7	8.69E+6	2.990E+6	0.7734	9.634E-5
210	1.84E+8	1.66E+7	4.455E+6	0.7773	5.447E-5
220	3.62E+8	3.16E+7	6.627E+6	0.7811	3.097E-5
230	7.19E+8	5.74E+7	9.644E+6	0.7845	1.770E-5
240	1.38E+9	1.09E+8	1.410E+7	0.7878	1.017E-5
250	2.62E+9	2.01E+8	2.049E+7	0.7911	5.876E-6
260	4.96E+9	3.66E+8	2.946E+7	0.7941	3.412E-6
270	9.29E+9	6.64E+8	4.204E+7	0.7969	1.992E-6
280	1.73E+10	1.19E+9	5.998E+7	0.7996	1.169E-6
290	3.16E+10	2.18E+9	8.558E+7	0.8023	6.888E-7
300	5.83E+10	3.97E+9	1.197E+8	0.8048	4.080E-7
310	1.06E+11	6.87E+9	1.678E+8	0.8072	2.428E-7
320	1.90E+11	1.20E+10	2.346E+8	0.8094	1.451E-7
330	3.38E+11	2.10E+10	3.297E+8	0.8117	8.713E-8

Note that the probability that an odd composite in the interval $(2^{k-1}, 2^k)$ passes the Fermat test is given by

$$P(2^{k-1}, 2^k) := \frac{S_c(2^{k-1}, 2^k)}{S_p(2^{k-1}, 2^k) + S_c(2^{k-1}, 2^k)}.$$

We have directly computed $S_p(2^{k-1}, 2^k)$ and $S_c(2^{k-1}, 2^k)$ for $k \leq 36$, with the latter computation aided by the formula $F(n) = \prod_{p|n} (p-1, n-1)$. Specifically, S_p is computed directly from the available list of primes up to 2^{36} . To compute S_c we use a sieve-like procedure. We initialize an array representing the odd numbers from 2^{k-1} and 2^k with all 1's. For each prime p to $2^k/3$, we let m run over the odd numbers between $2^{k-1}/p$ and $2^k/p$. For each m , we locate mp in the array, multiplying the entry there by $\gcd(m-1, p-1)$. At the end of the run the non-1 entries in our array correspond to the numbers $F(n)$ for n odd and composite. Note

this avoids factoring integers n in $(2^{k-1}, 2^k)$, though a brute force method to the modest level of 2^{36} would have worked too.

In Figure 4, we provide the values of $S_p(2^k)$ and $S_c(2^k)$, as well as $P(2^k)$ and $P(2^{k-1}, 2^k)$.

FIGURE 4. Exact values of data.

k	$S_p(2^{k-1}, 2^k)$	$S_c(2^{k-1}, 2^k)$	$P(2^{k-1}, 2^k)$	$P(2^k)$
3	6	0	0	0
4	18	2	$1.000E-1$	$7.692E-2$
5	104	4	$3.704E-2$	$4.478E-2$
6	320	24	$6.977E-2$	$6.276E-2$
7	1180	114	$8.810E-2$	$8.126E-2$
8	4292	316	$6.858E-2$	$7.210E-2$
9	16338	1114	$6.384E-2$	$6.605E-2$
10	57416	3056	$5.054E-2$	$5.492E-2$
11	208576	10890	$4.962E-2$	$5.109E-2$
12	780150	28094	$3.476E-2$	$3.922E-2$
13	2837158	74528	$2.600E-2$	$2.936E-2$
14	10673384	231514	$2.123E-2$	$2.342E-2$
15	39467286	582318	$1.454E-2$	$1.695E-2$
16	148222234	1636968	$1.092E-2$	$1.254E-2$
17	559288478	4521166	$8.019E-3$	$9.224E-3$
18	2106190104	11682336	$5.516E-3$	$6.503E-3$
19	7995006772	33290330	$4.147E-3$	$4.770E-3$
20	30299256236	88781082	$2.922E-3$	$3.410E-3$
21	115430158810	230250774	$1.991E-3$	$2.364E-3$
22	440353630422	628735800	$1.426E-3$	$1.672E-3$
23	1683364186642	1680806136	$9.975E-4$	$1.174E-3$
24	6448755473484	4408788648	$6.832E-4$	$8.115E-4$
25	24754014371036	11552686982	$4.665E-4$	$5.565E-4$
26	95132822935752	30756273488	$3.232E-4$	$3.840E-4$
27	366232744269106	82133627362	$2.242E-4$	$2.657E-4$
28	1411967930053822	215629423796	$1.527E-4$	$1.820E-4$
29	5450257882815404	565834872742	$1.038E-4$	$1.241E-4$
30	21065843780715212	1504267288346	$7.140E-5$	$8.504E-5$
31	81507897575948416	3999812059436	$4.907E-5$	$5.837E-5$
32	315718919767278610	10350692466866	$3.278E-5$	$3.940E-5$
33	1224166825030041460	27472503360964	$2.244E-5$	$2.682E-5$
34	4750936696054816476	72288538641772	$1.522E-5$	$1.821E-5$
35	18454541611019193346	190806759987694	$1.034E-5$	$1.237E-5$
36	71745407298862105164	498526567616818	$6.949E-6$	$8.342E-6$

Additionally, we have estimated $P(2^k)$ in the range $30 \leq k \leq 50$ using random sampling. More precisely, we randomly sample $\lfloor 2^{k/2} \rfloor$ odd composite numbers in

the interval $(2^{k-1}, 2^k)$, estimating $S_p(2^{k-1}, 2^k)$ by

$$\widehat{S}_p(2^{k-1}, 2^k) = \int_{2^{k-1}}^{2^k} \frac{t-3}{\log t} dt = \text{Li}(2^{2^k}) - \text{Li}(2^{2^{k-1}}) - 3(\text{Li}(2^k) - \text{Li}(2^{k-1})),$$

in order to smooth out some noise from the experiment. To estimate $S_c(2^{k-1}, 2^k)$, we add up $F(n) - 2$ for each odd composite n sampled, and scale this sum by

$$\frac{2^{k-2} - \text{Li}(2^k) + \text{Li}(2^{k-1})}{2^{k/2}},$$

representing the ratio between the number of composites in the interval and the number of samples taken. We repeat this procedure ten times, and compute the mean, $\widehat{S}_{\text{mean}}(2^{k-1}, 2^k)$, and median, $\widehat{S}_{\text{median}}(2^{k-1}, 2^k)$, of the data. Using these statistics, we estimate $P(2^{k-1}, 2^k)$ by

$$\begin{aligned} \widehat{P}_{\text{mean}}(2^{k-1}, 2^k) &= \frac{\widehat{S}_{\text{mean}}(2^{k-1}, 2^k)}{\widehat{S}_p(2^{k-1}, 2^k) + \widehat{S}_{\text{mean}}(2^{k-1}, 2^k)}, \\ \widehat{P}_{\text{median}}(2^{k-1}, 2^k) &= \frac{\widehat{S}_{\text{median}}(2^{k-1}, 2^k)}{\widehat{S}_p(2^{k-1}, 2^k) + \widehat{S}_{\text{median}}(2^{k-1}, 2^k)}. \end{aligned}$$

For $30 \leq k \leq 36$, $P(2^{k-1}, 2^k)$ is known, in which case we compute the relative errors, $\widehat{P}_{\text{mean}}/P - 1$ and $\widehat{P}_{\text{median}}/P - 1$, to get a sense of the accuracy of the experiment. Then we estimate $P(2^k)$ by

$$\widehat{P}_{\text{mean}}(2^k) = \frac{\widehat{S}_{\text{mean}}(2^k)}{\widehat{S}_p(2^k) + \widehat{S}_{\text{mean}}(2^k)}$$

where

$$\widehat{S}_{\text{mean}}(2^k) = \begin{cases} S_c(2^{k-1}) + \widehat{S}_{\text{mean}}(2^{k-1}, 2^k) & \text{for } 30 \leq k \leq 36, \\ S_c(2^{36}) + \sum_{j=37}^k \widehat{S}_{\text{mean}}(2^{j-1}, 2^j) & \text{for } 37 \leq k \leq 50, \end{cases}$$

and

$$\widehat{S}_p(2^k) = \begin{cases} S_p(2^{k-1}) + \widehat{S}_p(2^{k-1}, 2^k) & \text{for } 30 \leq k \leq 36, \\ S_p(2^{36}) + \sum_{j=37}^k \widehat{S}_p(2^{j-1}, 2^j) & \text{for } 37 \leq k \leq 50. \end{cases}$$

Results of the random sampling experiment are summarized in Figures 5 and 6.

One sees a negative bias in these data with the results of random sampling undershooting the true figures. The referee has pointed out to us that this may be due to Jensen's inequality applied to the convex function $x/(a+x)$, so that $E[X/(a+X)] \geq E[X]/(a+E[X])$. The undershoot may also be due to the fact that on average $F(n)$ is much larger than it is typically. In fact, it is shown in [8] that on a set of asymptotic density 1, we have $F(n) = n^{o(1)}$, yet the average behavior is $> n^{15/23}$. The exponent $15/23$, after more recent work of Baker and Harman [3], can be replaced with 0.7039. It follows from an old conjecture of Erdős on the distribution of Carmichael numbers that on average $F(n)$ behaves like $n^{1-o(1)}$.

FIGURE 5. Random sampling estimates in range where $P(2^k)$ is known.

k	$\widehat{P}_{\text{mean}}(2^{k-1}, 2^k)$	rel. err.	$\widehat{P}_{\text{median}}(2^{k-1}, 2^k)$	rel. err.	$\widehat{P}_{\text{mean}}(2^k)$	rel. err.
30	$5.541E-5$	-0.224	$5.045E-5$	-0.293	$7.319E-5$	-0.139
31	$4.800E-5$	-0.022	$3.616E-5$	-0.263	$5.758E-5$	-0.014
32	$2.706E-5$	-0.175	$1.899E-5$	-0.421	$3.515E-5$	-0.108
33	$2.223E-5$	-0.009	$1.248E-5$	-0.444	$2.666E-5$	-0.006
34	$1.387E-5$	-0.088	$1.013E-5$	-0.334	$1.721E-5$	-0.055
35	$7.603E-6$	-0.265	$6.506E-6$	-0.371	$1.033E-5$	-0.165
36	$4.433E-6$	-0.362	$4.123E-6$	-0.407	$6.474E-6$	-0.224

FIGURE 6. Random sampling estimates in range where $P(2^k)$ is unknown.

k	$\widehat{P}_{\text{mean}}(2^{k-1}, 2^k)$	$\widehat{P}_{\text{median}}(2^{k-1}, 2^k)$	$\widehat{P}_{\text{mean}}(2^k)$
37	$4.113E-6$	$3.675E-6$	$5.200E-6$
38	$4.807E-6$	$2.677E-6$	$4.908E-6$
39	$3.008E-6$	$1.463E-6$	$3.496E-6$
40	$1.519E-6$	$1.097E-6$	$2.026E-6$
41	$9.078E-7$	$5.697E-7$	$1.194E-6$
42	$7.747E-7$	$3.772E-7$	$8.822E-7$
43	$3.472E-7$	$2.334E-7$	$4.842E-7$
44	$1.968E-7$	$1.677E-7$	$2.704E-7$
45	$1.639E-7$	$1.687E-7$	$1.911E-7$
46	$1.186E-7$	$1.198E-7$	$1.372E-7$
47	$1.051E-7$	$6.597E-8$	$1.133E-7$
48	$4.076E-8$	$3.947E-8$	$5.928E-8$
49	$3.791E-8$	$3.213E-8$	$4.337E-8$
50	$2.361E-8$	$1.318E-8$	$2.865E-8$

ACKNOWLEDGMENTS

We gratefully acknowledge the many constructive comments of an anonymous referee. The first-named author is grateful for support received from the Byrne Scholars Program and the James O. Freedman Presidential Scholars program at Dartmouth College.

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. ———, *The difficulty of finding reliable witnesses*. Algorithmic Number Theory Proceedings (ANTS-I), L. M. Adleman and M.-D. Huang, eds., Lecture Notes in Computer Sci. **877** (1994), Springer-Verlag, Berlin, pp. 1–16.
3. R. C. Baker and G. Harman, *Shifted primes without large prime factors*. Acta Arith. **83** (1998), 331–361.
4. P. Beauchemin, G. Brassard, C. Crépeau, C. Goutier, and C. Pomerance, *The generation of random numbers that are probably prime*. J. Cryptology **1** (1988), 53–64.

5. R. J. Burthe, Jr., *Further investigations with the strong probable prime test*. Math. Comp. **65** (1996), 373–381.
6. J. Büthe, *Estimating $\pi(x)$ and related functions under partial RH assumptions*. Math. Comp. **85** (2016), 2483–2498.
7. ———, *An analytic method for bounding $\psi(x)$* . Math. Comp., to appear, see arXiv:1511.02032[math.NT].
8. P. Erdős, C. Pomerance, *On the number of false witnesses for a composite number*. Math. Comp. **46** (1986), 259–279.
9. I. Damgård, P. Landrock, C. Pomerance, *Average case error estimates for the strong probable prime test*. Math. Comp. **61** (1993), 177–194.
10. S. H. Kim, C. Pomerance, *The probability that a random probable prime is composite*. Math. Comp. **53** (1989), 721–741.
11. J. D. Lichtman and C. Pomerance, *Explicit estimates for the distribution of numbers free of large prime factors*. Submitted for publication, see arXiv:1705.02442 [math.NT].
12. L. Monier, *Evaluation and comparison of two efficient probabilistic primality testing algorithms*. Theoret. Comput. Sci. **12** (1980), 97–108.
13. D. J. Platt, T. S. Trudgian, *On the first sign change of $\vartheta(x) - x$* . Math. Comp. **85** (2016), 1539–1547.
14. M. O. Rabin, *Probabilistic algorithm for testing primality*. J. Number Theory, **12** (1980), 128–138.
15. R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.
16. J. Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers*. Illinois J. Math. **6** (1962), 64–94.
17. J. Rosser, L. Schoenfeld, *Sharper bounds for the Chebyshev functions $\vartheta(x)$ and $\psi(x)$* . Math. Comp. **29** (1975), 243–265.

DEPARTMENT OF MATHEMATICS, DARTMOUTH COLLEGE, HANOVER, NH 03755
E-mail address: lichtman.18@dartmouth.edu

DEPARTMENT OF MATHEMATICS, DARTMOUTH COLLEGE, HANOVER, NH 03755
E-mail address: carl.pomerance@dartmouth.edu