Supplemental Material for “A Statistical Comparison of Call Volume Uniformity Due to Mailing Strategy”

A. Simulations

We present several simulations to study properties of the procedures discussed in Section 3 in the article. Here, we consider a setting based on Scenario 1 and a setting based on 2, taking $J = 1$ in both. We present empirical rejection rates of hypothesis

$$H_0 : g(\boldsymbol{\theta}) \leq 0 \qquad \text{vs.} \quad H_1 : g(\boldsymbol{\theta}) > 0, \tag{A.1}$$

with significance level $\alpha = 0.05$, as well as the empirical coverage and a measure of width for level $1 - \alpha = 0.95$ lower confidence limits for $g(\boldsymbol{\theta})$.

A.1. Scenario 1

Consider the setting of Scenario 1 and suppose

$$p_1 \begin{pmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{15} \\ p_{16} \\ p_{17} \end{pmatrix} = \frac{1}{28} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{pmatrix} \approx \begin{pmatrix} 0.0357 \\ 0.0714 \\ 0.1071 \\ 0.1429 \\ 0.1786 \\ 0.2143 \\ 0.2500 \end{pmatrix} \text{ and } p_2 = \begin{pmatrix} p_{14} \\ p_{15} \\ p_{16} \\ p_{17} - \delta \\ p_{11} - \delta \\ p_{12} \\ p_{13} \end{pmatrix} \approx \begin{pmatrix} 0.1429 \\ 0.1786 \\ 0.2143 \\ 0.2500 - \delta \\ 0.0357 + \delta \\ 0.0714 \\ 0.1071 \end{pmatrix},$$

where $1/28$ is the normalizing constant needed to transform the vector $(1, 2, \ldots, 7)$ to a probability distribution. Here we consider the effect on $g(\boldsymbol{\theta})$ when $\boldsymbol{p}_2$ is a permutation of $\boldsymbol{p}_1$ except that the most frequent category donates some of its probability mass to the least frequent category. Restricting our attention to hypothesis (A.1), H_0 is true when $\delta = 0$ because $\boldsymbol{p}_1$ is exactly a permutation of $\boldsymbol{p}_2$. H_0 is also true for $\delta < 0$, where $\boldsymbol{p}_2$ becomes a more peaked distribution than $\boldsymbol{p}_1$. The value of $g(\boldsymbol{\theta}) = \mathcal{E}(\boldsymbol{p}_2) - \mathcal{E}(\boldsymbol{p}_1)$ increases to its maximum as δ increases from zero to 0.1071, but decreases as δ is increased further; see Figure 1a. The following steps are repeated 1,000 times:

1. Draw $\boldsymbol{X}_1 \sim \text{Mult}_k(m, \boldsymbol{p}_1)$ and $\boldsymbol{X}_2 \sim \text{Mult}_k(m, \boldsymbol{p}_2)$.
2. Compute the $\mathcal{Z}$-statistic and lower confidence limit $\mathcal{L}$ based on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

The empirical rejection rate is obtained from the proportion of rejections of the test $\mathcal{Z} > z_\alpha$. The empirical coverage of the lower confidence limit is obtained by the proportion of instances where

$\mathcal{L} \le g(\boldsymbol{\theta})$. The empirical width of the confidence limit is computed by averaging the individual widths defined by $g(\boldsymbol{\theta}) - \mathcal{L}$. This was repeated for each

$$\delta \in \{-0.02, -0.01, -0.005, -0.002, -0.001, -0.0005, -0.0002, -0.0001, 0, 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.14\},$$

and

$$m \in \{10, 20, 50, 100, 200, 500, 1000, 2000, 5000\}.$$

Tables 1, 2, and 3 present results for empirical rejection rate, empirical coverage, and empirical width respectively.

Results are mostly as expected, but some interesting features can be noted. Table 1 shows that the rejection rate reaches the nominal level of 0.05 as the sample size increases and δ approaches 0 from below. The power of the test increases when δ approaches the value 0.1071, which maximizes $g(\boldsymbol{\theta})$, with the increase being faster as the sample size is taken larger. We notice some oscillations in the power; for example, when $\delta = 0$, the rejection rate reduces from 0.0490 at $m = 2000$ to 0.0460 at 5000. Table 2 shows that coverage probability for the confidence limit approaches the nominal 0.95 level as sample size becomes large, for all values of δ. Oscillations in coverage probability can be seen, for example, when $\delta = 0.0002$ and m increases from 500 to 5000. Empirical width shown in Table 3 appears to be decreasing for all δ as sample size is increased. However, for fixed δ widths appear to become smaller as $g(\boldsymbol{\theta})$ is taken larger.

Regarding the oscillations, our $\mathcal{Z}$-statistic is closely related to the Wald statistic used for inference on p in the situation $X \sim \text{Binomial}(m, p)$. Brown et al. (2001) discuss at length how coverage probabilities for confidence intervals based on the Wald statistic exhibit an oscillating behavioras m and p vary. For example, with p fixed, coverage probability will meet the nominal level for some m and fail to meet the nominal level for some larger m. Brown et al. (2001) suggest several alternative intervals based on the binomial data, and more recently Franco et al. (2019) and Kott (2017) compare binomial intervals in the setting of complex surveys. Alternative intervals couldalso be considered in our setting.

A.2. Scenario 2

Now consider the setting of Scenario 2 and let

$$\boldsymbol{p}_1 = \frac{1}{28}\begin{pmatrix}1\\2\\3\\4\\5\\6\\7\end{pmatrix} \approx \begin{pmatrix}0.0357\\0.0714\\0.1071\\0.1429\\0.1786\\0.2143\\0.2500\end{pmatrix}$$

$\boldsymbol{q} = \pi \boldsymbol{p}_2 + (1-\pi)\,\boldsymbol{p}_3$ with $\pi = 1/2$, $\boldsymbol{p}_2 = \boldsymbol{p}_1$, and

$$\boldsymbol{p}_3 = (1-\delta)\boldsymbol{p}_1 + \delta\frac{1}{28}\begin{pmatrix}7\\6\\5\\4\\3\\2\\1\end{pmatrix} = \frac{1}{28}\begin{pmatrix}(1-\delta)1+\delta 7\\(1-\delta)2+\delta 6\\(1-\delta)3+\delta 5\\(1-\delta)4+\delta 4\\(1-\delta)5+\delta 3\\(1-\delta)6+\delta 2\\(1-\delta)7+\delta 1\end{pmatrix}$$

Focusing on hypothesis (A.1), H_0 is true when $\delta = 0$ so that $\boldsymbol{p}_1 = \boldsymbol{q}$. The value of $g(\boldsymbol{\theta}) = \mathcal{E}(\boldsymbol{q}) - \mathcal{E}(\boldsymbol{p})$ increases to its maximum when $\delta = 1$, where

$$\boldsymbol{q} = \frac{1}{28}[\pi(1, \dots, 7) + (1 - \pi)(7, \dots, 1)] = \frac{1}{7}(1, \dots, 1)$$

is the discrete uniform distribution. Figure 1b displays a plot of $g(\boldsymbol{\theta})$ for $\delta \in (0, 1)$. The followingsteps are repeated 1,000 times:

1. Draw $m_2 \sim \text{Binomial}(m, \pi)$ and let $m_3 = m - m_2$.
2. Draw $\boldsymbol{X}_1 \sim \text{Mult}_{\boldsymbol{k}}(m, \boldsymbol{p}_1)$, $\boldsymbol{X}_2 \sim \text{Mult}_{\boldsymbol{k}}(m_2, \boldsymbol{p}_2)$, and $\boldsymbol{X}_3 \sim \text{Mult}_{\boldsymbol{k}}(m_3, \boldsymbol{p}_3)$.
3. Compute the $\mathcal{Z}$-statistic and lower confidence limit $\mathcal{L}$ based on $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\hat{\pi} = m_2/(m_2 + m_3)$.

Scenario 2 is otherwise similar to Scenario 1, except that we consider

$$\delta \in \{0, 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$$
$$\text{and} \quad m \in \{50, 100, 200, 500, 1000, 2000, 5000\}.$$

We avoided the smallest sample sizes from Scenario 1 which occasionally resulted in $m_2 = 0$ or $m_3 = 0$.

Tables 4, 5, and 6 present results for empirical rejection rate, empirical coverage, and empirical width respectively. As in Scenario 1, results here are mostly as expected, with similar oscillations in rejection rate and coverage. The effect size $g(\boldsymbol{\theta})$ becomes substantially larger here than in Scenario1 for the larger choices of δ; see Figure 1b.

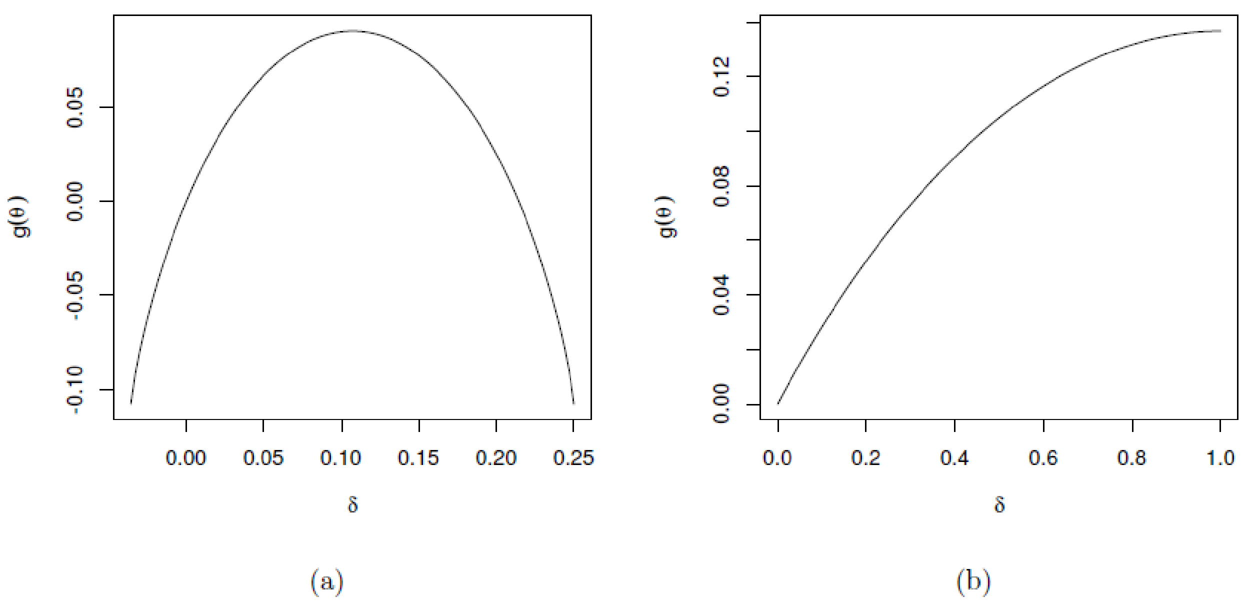

Fig. 1. The value of $g(\boldsymbol{\theta})$ for: (a) $\delta \in (-0.04, 0.25)$, when $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are as described in Subsection A.1, and (b) $\delta \in (0,1)$, when $\boldsymbol{p}_1$ and $\boldsymbol{q}$ are as described in Subsection A.2.

Table 1. Empirical rejection rate of one-sided test from simulation study for scenario 1.

δ	$g(\boldsymbol{\theta})$	$m = 10$	20	50	100	200	500	1,000	2,000	5,000
-0.0200	-0.0468	0.1340	0.0850	0.0460	0.0150	0.0050	0.0000	0.0000	0.0000	0.0000
-0.0100	-0.0212	0.1600	0.1070	0.0480	0.0290	0.0120	0.0070	0.0060	0.0010	0.0000
-0.0050	-0.0101	0.1590	0.1180	0.0690	0.0380	0.0310	0.0190	0.0160	0.0100	0.0020
-0.0020	-0.0040	0.1930	0.1270	0.0760	0.0550	0.0390	0.0350	0.0280	0.0330	0.0180
-0.0010	-0.0020	0.1620	0.1350	0.0780	0.0530	0.0470	0.0380	0.0370	0.0380	0.0360
-0.0005	-0.0010	0.1920	0.1140	0.0760	0.0400	0.0420	0.0450	0.0430	0.0450	0.0470
-0.0002	-0.0004	0.1620	0.1080	0.0640	0.0600	0.0320	0.0370	0.0550	0.0450	0.0390
-0.0001	-0.0002	0.1410	0.1160	0.0800	0.0450	0.0490	0.0380	0.0440	0.0530	0.0440
0.0000	0.0000	0.1640	0.1100	0.0800	0.0510	0.0450	0.0550	0.0430	0.0490	0.0460
0.0001	0.0002	0.1810	0.1250	0.0890	0.0600	0.0490	0.0570	0.0430	0.0490	0.0670
0.0002	0.0004	0.1730	0.1070	0.0680	0.0440	0.0490	0.0430	0.0370	0.0560	0.0640
0.0005	0.0010	0.1740	0.1140	0.0740	0.0660	0.0680	0.0520	0.0510	0.0520	0.0630
0.0010	0.0019	0.1620	0.1450	0.0870	0.0560	0.0510	0.0550	0.0530	0.0600	0.0710
0.0020	0.0038	0.1930	0.1450	0.0820	0.0650	0.0630	0.0660	0.0680	0.0840	0.1080
0.0050	0.0093	0.1690	0.1290	0.0860	0.0720	0.0680	0.0810	0.1300	0.1650	0.2810
0.0100	0.0180	0.1900	0.1460	0.0950	0.0820	0.0810	0.1630	0.2160	0.3580	0.6180
0.0200	0.0333	0.1840	0.1580	0.1390	0.1390	0.1730	0.3260	0.5200	0.7870	0.9850
0.0500	0.0669	0.2340	0.1910	0.2030	0.2490	0.4810	0.8170	0.9780	0.9990	1.0000
0.1000	0.0900	0.2600	0.2460	0.2520	0.4050	0.7230	0.9770	0.9990	1.0000	1.0000
0.1400	0.0828	0.2540	0.2120	0.2560	0.3800	0.6110	0.9530	0.9980	1.0000	1.0000

Table 2. Empirical coverage rate of lower confidence limit from simulation study for scenario 1.

δ	$g(\boldsymbol{\theta})$	$m = 10$	20	50	100	200	500	1,000	2,000	5,000
-0.0200	-0.0468	0.8220	0.8760	0.9170	0.9440	0.9530	0.9500	0.9430	0.9430	0.9450
-0.0100	-0.0212	0.8280	0.8770	0.9340	0.9330	0.9440	0.9580	0.9520	0.9450	0.9530
-0.0050	-0.0101	0.8350	0.8730	0.9180	0.9450	0.9360	0.9580	0.9510	0.9520	0.9440
-0.0020	-0.0040	0.8050	0.8700	0.9220	0.9410	0.9520	0.9440	0.9570	0.9500	0.9540
-0.0010	-0.0020	0.8380	0.8630	0.9180	0.9430	0.9470	0.9550	0.9570	0.9540	0.9480
-0.0005	-0.0010	0.8080	0.8840	0.9230	0.9580	0.9550	0.9520	0.9560	0.9470	0.9440
-0.0002	-0.0004	0.8380	0.8920	0.9350	0.9390	0.9680	0.9620	0.9420	0.9530	0.9600
-0.0001	-0.0002	0.8590	0.8840	0.9200	0.9550	0.9510	0.9620	0.9540	0.9470	0.9550
0.0000	0.0000	0.8360	0.8900	0.9200	0.9490	0.9550	0.9450	0.9570	0.9510	0.9540
0.0001	0.0002	0.8190	0.8750	0.9120	0.9400	0.9510	0.9440	0.9580	0.9540	0.9380
0.0002	0.0004	0.8270	0.8930	0.9320	0.9570	0.9520	0.9580	0.9660	0.9480	0.9400
0.0005	0.0010	0.8260	0.8890	0.9270	0.9350	0.9330	0.9500	0.9540	0.9540	0.9500
0.0010	0.0019	0.8380	0.8560	0.9140	0.9460	0.9550	0.9530	0.9540	0.9520	0.9520
0.0020	0.0038	0.8090	0.8580	0.9210	0.9360	0.9460	0.9500	0.9530	0.9500	0.9500
0.0050	0.0093	0.8410	0.8790	0.9230	0.9450	0.9590	0.9580	0.9470	0.9490	0.9460
0.0100	0.0180	0.8330	0.8730	0.9270	0.9450	0.9610	0.9420	0.9520	0.9480	0.9490
0.0200	0.0333	0.8430	0.8780	0.9080	0.9430	0.9530	0.9460	0.9480	0.9580	0.9540
0.0500	0.0669	0.8320	0.8860	0.9080	0.9400	0.9530	0.9550	0.9660	0.9510	0.9580
0.1000	0.0900	0.8310	0.8680	0.9190	0.9580	0.9630	0.9490	0.9450	0.9490	0.9480
0.1400	0.0828	0.8330	0.8830	0.9180	0.9480	0.9410	0.9510	0.9530	0.9480	0.9500

Table 3. Empirical width of lower confidence limit from simulation study for scenario 1.

δ	$g(\boldsymbol{\theta})$	$m = 10$	20	50	100	200	500	1,000	2,000	5,000
-0.0200	-0.0468	0.2639	0.2319	0.1647	0.1142	0.0817	0.0508	0.0358	0.0246	0.0156
-0.0100	-0.0212	0.2634	0.2417	0.1711	0.1152	0.0786	0.0503	0.0357	0.0242	0.0156
-0.0050	-0.0101	0.2706	0.2345	0.1590	0.1175	0.0795	0.0499	0.0353	0.0248	0.0151
-0.0020	-0.0040	0.2697	0.2302	0.1669	0.1166	0.0808	0.0484	0.0349	0.0242	0.0157
-0.0010	-0.0020	0.2773	0.2291	0.1629	0.1152	0.0787	0.0490	0.0347	0.0245	0.0153
-0.0005	-0.0010	0.2490	0.2516	0.1635	0.1196	0.0790	0.0512	0.0345	0.0247	0.0148
-0.0002	-0.0004	0.2841	0.2428	0.1655	0.1124	0.0809	0.0492	0.0357	0.0246	0.0158
-0.0001	-0.0002	0.2898	0.2485	0.1670	0.1116	0.0793	0.0491	0.0341	0.0233	0.0149
0.0000	0.0000	0.2903	0.2380	0.1629	0.1132	0.0806	0.0498	0.0345	0.0243	0.0155
0.0001	0.0002	0.2737	0.2438	0.1613	0.1106	0.0771	0.0475	0.0350	0.0235	0.0149
0.0002	0.0004	0.2757	0.2449	0.1684	0.1137	0.0802	0.0498	0.0337	0.0237	0.0152
0.0005	0.0010	0.2748	0.2416	0.1660	0.1110	0.0787	0.0483	0.0350	0.0239	0.0155
0.0010	0.0019	0.2928	0.2269	0.1629	0.1129	0.0812	0.0482	0.0345	0.0244	0.0151
0.0020	0.0038	0.2613	0.2370	0.1625	0.1152	0.0791	0.0480	0.0337	0.0243	0.0157
0.0050	0.0093	0.2782	0.2428	0.1625	0.1115	0.0789	0.0496	0.0336	0.0238	0.0150
0.0100	0.0180	0.2733	0.2334	0.1630	0.1136	0.0788	0.0471	0.0331	0.0239	0.0152
0.0200	0.0333	0.2945	0.2372	0.1513	0.1110	0.0752	0.0461	0.0334	0.0226	0.0145
0.0500	0.0669	0.2868	0.2383	0.1527	0.1079	0.0706	0.0442	0.0312	0.0217	0.0136
0.1000	0.0900	0.2776	0.2253	0.1468	0.1021	0.0676	0.0421	0.0286	0.0208	0.0128
0.1400	0.0828	0.2787	0.2392	0.1480	0.1025	0.0690	0.0422	0.0301	0.0208	0.0133

Table 4. Empirical rejection rate of one-sided test from simulation study for scenario 2.

δ	$g(\boldsymbol{\theta})$	$m = 50$	100	200	500	1,000	2,000	5,000
0.0000	0.0000	0.0800	0.0540	0.0430	0.0520	0.0390	0.0470	0.0480
0.0001	0.0000	0.0700	0.0620	0.0380	0.0620	0.0580	0.0430	0.0590
0.0002	0.0001	0.0780	0.0610	0.0470	0.0550	0.0390	0.0590	0.0590
0.0005	0.0002	0.0800	0.0680	0.0540	0.0560	0.0660	0.0570	0.0520
0.0010	0.0003	0.0880	0.0470	0.0430	0.0520	0.0410	0.0410	0.0630
0.0020	0.0006	0.0680	0.0540	0.0520	0.0520	0.0560	0.0590	0.0480
0.0050	0.0015	0.0710	0.0560	0.0550	0.0710	0.0620	0.0590	0.0690
0.0100	0.0030	0.0880	0.0690	0.0650	0.0680	0.0650	0.0710	0.0880
0.0200	0.0060	0.0790	0.0530	0.0750	0.0750	0.0880	0.1120	0.1630
0.0500	0.0147	0.0930	0.0760	0.0940	0.1250	0.2020	0.2430	0.5090
0.1000	0.0283	0.1200	0.1140	0.1490	0.2560	0.4370	0.6140	0.9410
0.2000	0.0525	0.1680	0.2200	0.2920	0.6340	0.8620	0.9910	1.0000
0.5000	0.1049	0.3200	0.5500	0.8560	0.9970	1.0000	1.0000	1.0000
1.0000	0.1368	0.4520	0.8350	0.9970	1.0000	1.0000	1.0000	1.0000

Table 5. *Empirical coverage of lower confidence limit from simulation study for scenario 2.*

δ	$g(\boldsymbol{\theta})$	$m = 50$	100	200	500	1000	2000	5000
0.0000	0.0000	0.9200	0.9460	0.9570	0.9480	0.9610	0.9530	0.9520
0.0001	0.0000	0.9300	0.9380	0.9620	0.9380	0.9420	0.9580	0.9410
0.0002	0.0001	0.9220	0.9390	0.9540	0.9450	0.9620	0.9410	0.9420
0.0005	0.0002	0.9200	0.9320	0.9460	0.9460	0.9350	0.9460	0.9500
0.0010	0.0003	0.9130	0.9530	0.9580	0.9480	0.9600	0.9600	0.9400
0.0020	0.0006	0.9320	0.9460	0.9510	0.9500	0.9450	0.9450	0.9590
0.0050	0.0015	0.9310	0.9460	0.9490	0.9370	0.9480	0.9530	0.9460
0.0100	0.0030	0.9150	0.9370	0.9480	0.9440	0.9520	0.9600	0.9500
0.0200	0.0060	0.9290	0.9570	0.9360	0.9470	0.9440	0.9490	0.9490
0.0500	0.0147	0.9240	0.9500	0.9440	0.9580	0.9490	0.9500	0.9360
0.1000	0.0283	0.9200	0.9550	0.9450	0.9580	0.9430	0.9570	0.9430
0.2000	0.0525	0.9130	0.9430	0.9510	0.9490	0.9580	0.9520	0.9390
0.5000	0.1049	0.9180	0.9470	0.9640	0.9550	0.9560	0.9450	0.9370
1.0000	0.1368	0.9120	0.9560	0.9550	0.9610	0.9560	0.9650	0.9500

Table 6. *Empirical width of lower confidence limit from simulation study for scenario 2.*

δ	$g(\boldsymbol{\theta})$	$m = 50$	100	200	500	1,000	2,000	5,000
0.0000	0.0000	0.1556	0.1133	0.0791	0.0471	0.0353	0.0244	0.0155
0.0001	0.0000	0.1615	0.1129	0.0794	0.0472	0.0337	0.0249	0.0153
0.0002	0.0001	0.1572	0.1098	0.0782	0.0487	0.0360	0.0234	0.0153
0.0005	0.0002	0.1592	0.1160	0.0783	0.0473	0.0341	0.0245	0.0152
0.0010	0.0003	0.1562	0.1134	0.0786	0.0488	0.0354	0.0251	0.0153
0.0020	0.0006	0.1591	0.1135	0.0788	0.0493	0.0340	0.0240	0.0158
0.0050	0.0015	0.1615	0.1159	0.0807	0.0474	0.0341	0.0245	0.0150
0.0100	0.0030	0.1595	0.1122	0.0793	0.0494	0.0346	0.0244	0.0151
0.0200	0.0060	0.1621	0.1164	0.0773	0.0489	0.0344	0.0248	0.0150
0.0500	0.0147	0.1597	0.1115	0.0784	0.0487	0.0325	0.0242	0.0146
0.1000	0.0283	0.1547	0.1137	0.0769	0.0470	0.0321	0.0238	0.0145
0.2000	0.0525	0.1499	0.1043	0.0761	0.0441	0.0316	0.0215	0.0141
0.5000	0.1049	0.1401	0.0961	0.0656	0.0393	0.0271	0.0185	0.0119
1.0000	0.1368	0.1341	0.0915	0.0580	0.0358	0.0250	0.0171	0.0111

B. References

Brown, L. D., T. T. Cai, and A. DasGupta, 2001. “Interval Estimation for a Binomial Proportion”. *Statistical Science*, 16: 101–117. DOI: https://dx.doi.org/10.1214/ss/1009213286.

Franco, C., R.J.A. Little, T.A. Louis, and E.V. Slud, 2019. "Comparative Study of Confidence Intervals for Proportions in Complex Sample Surveys". *Journal of Survey Statistics and Methodology*, 7: 334–364. DOI: https://doi.org/10.1093/jssam/smy019.

Kott, P. S. 2017. “A note on Wilson coverage intervals for proportions estimated from complex samples”. *Survey Methodology*, 43: 235–240. Available at: https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201700254872