Supplementary Materials: Gini Index Estimation Within Pre-specified Error Bound: Application to Indian Household Survey Data

Francis Bilson Darku [1,†], Frank Konietschke [2] and Bhargab Chattopadhyay [3,*]

We use simulation and replication studies to illustrate and compare the properties of our purely sequential and the two stage procedures in constructing a $100(1-\alpha)\%$ confidence interval for the Gini index whose width is less than ω, under a complex household survey. We present two different simulation studies with same simulation sizes - (a) simulation using the NSS survey data as the population and (b) a Monte Carlo simulation in which the observations are drawn from three populations, each of which has been drawn using three different distributions, namely; Pareto, Gamma and Lognormal distributions.

1. Replication Using Empirical Data

1.1. Replication Using Empirical Data: Purely Sequential Procedure (PSP)

The purely sequential procedure was replicated 5000 times on the NSS data. The results of the simulation study of our purely sequential procedure are found in Tables S1 to S4. In the tables, the average estimated cluster sizes $\bar{N}$ and their respective standard errors $se(N)$ are indicated in the fifth column. The 7th column, shows the coverage probability p of the confidence intervals and its standard error s_p, assuming $\hat{G}_H$ to be the true population Gini index. In the 9th column, $\Pr(w_N > \omega)$ indicates the proportion of confidence interval widths that exceeded the desired bound ω, while in the last column, $\Pr(N < \hat{C})$ is the proportion of estimated optimal cluster sizes that could not meet the stopping rule.

Apart from the urban sectors, the average confidence interval width $\bar{w}_N$ are less than $\omega = 0.02$ and they have small standard errors. Also, none of the width exceeded the specified value of ω. The coverage probability under each scenario was not below $(1-\alpha)\%$. It can be noted that the coverage probabilities for the Urban regions were mostly 1. These are as a result of the wider confidence intervals that were obtained at the end of the procedure.

1.2. Replication Using Empirical Data: Two-Stage procedure

Tables S5 to S8 show the results of the simulation study of our two-stage procedure. The third column shows the average estimated Gini Index and its standard error. The fourth through to the sixth columns show the average values of Q^*, Q, and $\tilde{Q}$ respectively with their standard errors. The seventh column provides the coverage probabilities p (assuming $\hat{G}_H$ to be the true population value) and its standard error. In the eighth column is the average confidence interval width $w_{\tilde{Q}}$ and its standard error. The last column indicates the proportion of times that the length of the confidence intervals obtained ($w_{\tilde{Q}}$) exceeded the pre-specified upper bound ω.

As the application of the two-stage procedure was repeated 5000, results show that the optimal cluster size is less than the number of clusters in the NSS data on the average, independent of the households that were in the pilot sample. However, the estimated optimal cluster sizes (Q^*) in the urban clusters always exceeded the available number of clusters irrespective of which clusters and households were initially sampled. This is clearly seen in the last column of Tables S5 to S8. The urban sectors always had confidence interval widths more than desired due to inadequate number of clusters. Unsurprisingly, these wider confidence intervals resulted in coverage probabilities of 1. Also, from the penultimate column, it could be seen that there are chances that this procedure result in a larger width.

Table S1. Replication results for PSP on 64th NSS Data ($\alpha = 0.10$, $\omega = 0.02$)

Region	$\hat{G}_H$	H	$\overline{\hat{C}}$	$\bar{N}$	$\overline{\hat{G}}_N$	p	$\bar{w}_N$	$\Pr(w_N > \omega)$	$\Pr(N < \hat{C})$
	$se(\hat{G}_H)$	t	$se(\hat{C})$	$se(N)$	$se(\hat{G}_N)$	s_p	$se(\bar{w}_N)$	$se(\Pr(w_N > \omega))$	
Uttar Pradesh									
All	0.2163	1262	597.3592	637.7388	0.2160	0.9672	0.0186	0.0000	0.0000
	0.0042	321	112.5162	117.8815	0.0042	0.0025	0.0006	0.0000	
Rural	0.1997	903	431.5098	450.9542	0.1993	0.9658	0.0180	0.0000	0.0000
	0.0041	198	77.5388	78.3839	0.0041	0.0012	0.0007	0.0000	
Urban	0.2229	359	903.0000	359.0000	0.2229	1.0000	0.0304	1.0000	1.0000
	0.0092	180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
West Bengal									
All	0.2320	878	682.2386	691.2456	0.2315	0.9772	0.0193	0.0000	0.0000
	0.0051	190	85.2230	85.8925	0.0036	0.0021	0.0002	0.0000	
Rural	0.1812	551	391.9578	397.1726	0.1808	0.9804	0.0179	0.0000	0.0000
	0.0048	172	51.0689	52.2322	0.0032	0.0020	0.0006	0.0000	
Urban	0.2609	327	612.0000	327.0000	0.2609	1.0000	0.0254	1.0000	1.0000
	0.0077	185	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Table S2. Replication results for PSP on 64th NSS Data ($\alpha = 0.05$, $\omega = 0.02$)

Region	$\hat{G}_H$	H	$\overline{\hat{C}}$	$\bar{N}$	$\overline{\hat{G}}_N$	p	$\bar{w}_N$	$\Pr(w_N > \omega)$	$\Pr(N < \hat{C})$
	$se(\hat{G}_H)$	t	$se(\hat{C})$	$se(N)$	$se(\hat{G}_N)$	s_p	$se(\bar{w}_N)$	$se(\Pr(w_N > \omega))$	
Uttar Pradesh									
All	0.2163	1262	836.5332	871.9902	0.2159	0.9904	0.0190	0.0000	0.0000
	0.0042	321	111.8031	113.5042	0.0030	0.0014	0.0004	0.0000	
Rural	0.1997	903	592.8814	612.3790	0.1992	0.9930	0.0185	0.0000	0.0000
	0.0041	198	88.4435	89.3062	0.0029	0.0012	0.0005	0.0000	
Urban	0.2229	359	1282.0000	359.0000	0.2229	1.0000	0.0362	1.0000	1.0000
	0.0092	180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
West Bengal									
All	0.2320	878	888.4526	866.3402	0.2316	0.9980	0.0197	0.0000	0.8218
	0.0051	190	42.1969	31.2673	0.0013	0.0006	0.0002	0.0000	
Rural	0.1812	551	531.0336	530.8254	0.1807	0.9978	0.0186	0.0000	0.7352
	0.0048	172	42.5598	41.2625	0.0014	0.0007	0.0003	0.0000	
Urban	0.2609	327	869.0000	327.0000	0.2609	1.0000	0.0303	1.0000	1.0000
	0.0077	185	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Table S3. Replication results for PSP on 64th NSS Data ($\alpha = 0.10$, $\omega = 0.025$)

Region	$\hat{G}_H$	H	$\overline{\hat{C}}$	$\bar{N}$	$\overline{\hat{G}}_N$	p	$\bar{w}_N$	$\Pr(w_N > \omega)$	$\Pr(N < \hat{C})$
	$se(\hat{G}_H)$	t	$se(\hat{C})$	$se(N)$	$se(\hat{G}_N)$	s_p	$se(\bar{w}_N)$	$se(\Pr(w_N > \omega))$	
Uttar Pradesh									
All	0.2163	1262	397.4636	446.5638	0.2162	0.9400	0.0222	0.0000	0.0000
	0.0042	321	87.6632	86.6652	0.0058	0.0034	0.0012	0.0000	
Rural	0.1997	903	290.1868	323.1846	0.1995	0.9438	0.0211	0.0000	0.0000
	0.0041	168	61.1267	66.2810	0.0054	0.0033	0.0012	0.0000	
Urban	0.2229	359	578.0000	359.0000	0.2229	1.0000	0.0304	1.0000	1.0000
	0.0092	180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
West Bengal									
All	0.2320	878	458.0644	468.1314	0.2315	0.9528	0.0235	0.0000	0.0000
	0.0051	190	92.1715	93.6177	0.0057	0.0030	0.0006	0.0000	
Rural	0.1812	551	268.8992	274.2660	0.1812	0.9624	0.0213	0.0000	0.0000
	0.0048	172	44.5474	45.7673	0.0049	0.0027	0.0013	0.0000	
Urban	0.2609	327	389.7278	326.3296	0.2607	0.9964	0.0254	0.9760	0.9760
	0.0077	185	14.7339	5.0234	0.0005	0.0008	0.1531	0.0022	

Table S4. Replication results for PSP on 64th NSS Data ($\alpha = 0.05, \omega = 0.025$)

Region	$\hat{G}_H$ / $se(\hat{G}_H)$	H / t	$\overline{\hat{C}}$ / $se(\hat{C})$	$\bar{N}$ / $se(N)$	$\overline{\hat{G}}_N$ / $se(\hat{G}_N)$	p / s_p	$\bar{w}_N$ / $se(\bar{w}_N)$	$\Pr(w_N > \omega)$ / $se(\Pr(w_N > \omega))$	$\Pr(N < \hat{C})$
Uttar Pradesh									
All	0.2163	1262	549.4374	588.3550	0.2161	0.9796	0.0231	0.0000	0.0000
	0.0042	321	107.3979	110.8163	0.0046	0.0020	0.0008	0.0000	
Rural	0.1997	903	399.6852	417.5508	0.1994	0.9814	0.0223	0.0000	0.0000
	0.0041	198	78.7000	75.5232	0.0043	0.0019	0.0010	0.0000	
Urban	0.2229	359	821.0000	359.0000	0.2229	1.0000	0.0362	1.0000	1.0000
	0.0092	180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
West Bengal									
All	0.2320	878	632.3722	641.8234	0.2315	0.9854	0.0240	0.0000	0.0000
	0.0051	190	92.1456	93.1628	0.0040	0.0017	0.0004	0.0000	
Rural	0.1812	551	361.3896	366.2682	0.1809	0.9868	0.0222	0.0000	0.0000
	0.0048	172	50.9696	52.3207	0.0036	0.0016	0.0009	0.0000	
Urban	0.2609	327	556.0000	327.0000	0.2609	1.0000	0.0303	1.0000	1.0000
	0.0077	185	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Table S5. Replication results for two-stage procedure on 64th NSS Data ($\alpha = 0.1, \omega = 0.02$)

Region	$\hat{G}_H$ / $se(\hat{G}_H)$	H / t	$\overline{\hat{G}}_{\tilde{Q}}$ / $se(\hat{G}_{\tilde{Q}})$	$\overline{Q^*}$ / $se(Q^*)$	$\overline{Q}$ / $se(Q)$	$\overline{\tilde{Q}}$ / $se(\tilde{Q})$	p / s_p	$\overline{w}_{\tilde{Q}}$ / $se(w_{\tilde{Q}})$	$\Pr(w_{\tilde{Q}} > \omega)$ / $se(\Pr(w_{\tilde{Q}} > \omega))$
Utter Pradesh									
All	0.2163	1262	0.2158	746.0006	736.7442	747.7408	0.9664	0.0175	0.1940
	0.0042	321	0.0038	292.0440	270.7270	260.7141	0.0025	0.0028	0.0059
Rural	0.1997	903	0.1993	521.3184	521.0958	521.2806	0.9648	0.0173	0.2060
	0.0041	198	0.0038	173.9302	173.4148	173.5880	0.0026	0.0029	0.0057
Urban	0.2229	359	0.2229	921.8320	358.9994	359.0000	1.0000	0.0304	1.0000
	0.0092	180	0.0000	234.4798	0.0424	0.0000	0.0000	0.0000	0.0000
West Bengal									
All	0.2320	878	0.2317	783.1276	719.0094	719.0098	0.9874	0.0184	0.2248
	0.0051	190	0.0029	240.1110	151.8134	152.2045	0.0016	0.0021	0.0059
Rural	0.1812	551	0.1809	491.3992	459.2752	459.2734	0.9886	0.0170	0.0706
	0.0048	173	0.0024	123.0809	78.5944	78.6132	0.0015	0.0017	0.0036
Urban	0.2609	327	0.2609	696.5434	327.0000	327.0000	1.0000	0.0254	1.0000
	0.0077	185	0.0000	113.4849	0.0000	0.0000	0.0000	0.0000	0.0000

Table S6. Replication results for two-stage procedure on 64th NSS Data ($\alpha = 0.05, \omega = 0.02$)

Region	$\hat{G}_H$ / $se(\hat{G}_H)$	H / t	$\overline{\hat{G}}_{\tilde{Q}}$ / $se(\hat{G}_{\tilde{Q}})$	$\overline{Q^*}$ / $se(Q^*)$	$\overline{Q}$ / $se(Q)$	$\overline{\tilde{Q}}$ / $se(\tilde{Q})$	p / s_p	$\overline{w}_{\tilde{Q}}$ / $se(w_{\tilde{Q}})$	$\Pr(w_{\tilde{Q}} > \omega)$ / $se(\Pr(w_{\tilde{Q}} > \omega))$
Uttar Pradesh									
All	0.2163	1262	0.2160	1050.3342	938.7144	937.7854	0.9910	0.0189	0.3372
	0.0042	333	0.0028	427.1341	253.3348	255.7325	0.0013	0.0026	0.0067
Rural	0.1997	903	0.1995	737.4964	660.4210	660.4842	0.9948	0.0185	0.3296
	0.0041	226	0.0027	280.1239	147.8075	147.9787	0.0010	0.0026	0.0066
Urban	0.2229	359	0.2229	1438.5742	359.0000	359.0000	1.0000	0.0362	1.0000
	0.0092	254	0.0000	324.5592	0.0000	0.0000	0.0000	0.0000	0.0000
West Bengal									
All	0.2320	878	0.2319	1067.1052	833.7042	833.7296	0.9998	0.0203	0.2620
	0.0051	223	0.0014	292.5594	78.2840	78.5031	0.0002	0.0011	0.0062
Rural	0.1812	551	0.1811	666.8132	534.2016	534.2286	1.0000	0.0189	0.1012
	0.0048	203	0.0010	160.5609	31.3909	31.4040	0.0000	0.0007	0.0043
Urban	0.2609	327	0.2609	950.6524	327.0000	327.0000	1.0000	0.0303	1.0000
	0.0077	207	0.0000	132.2993	0.0000	0.0000	0.0000	0.0000	0.0000

Table S7. Replication results for two-stage procedure on 64th NSS Data ($\alpha = 10\%, \omega = 0.025, k \leq 4$)

Region	$\hat{G}_H$ / $se(\hat{G}_H)$	H / t	$\overline{\hat{G}}_{\tilde{Q}}$ / $se(\hat{G}_{\tilde{Q}})$	$\overline{Q^*}$ / $se(Q^*)$	$\overline{Q}$ / $se(Q)$	$\overline{\tilde{Q}}$ / $se(\tilde{Q})$	p / s_p	$\overline{w}_{\tilde{Q}}$ / $se(w_{\tilde{Q}})$	$p(w_{\tilde{Q}} > \omega)$ / $se(p(w_{\tilde{Q}} > \omega))$
Uttar Pradesh									
All	0.2163	1262	0.2157	511.2518	511.2518	540.5866	0.9334	0.0199	0.0374
	0.0042	302	0.0052	174.7202	174.7202	151.6189	0.0035	0.0026	0.0027
Rural	0.1997	903	0.1994	374.6068	374.5782	374.7222	0.9464	0.0204	0.1462
	0.0041	168	0.0053	176.9589	176.8717	177.1875	0.0032	0.0043	0.0050
Urban	0.2229	359	0.2228	658.1510	357.3200	357.6046	0.9990	0.0304	0.9980
	0.0092	168	0.0009	227.5936	8.9762	9.9289	0.0004	0.0005	0.0000
West Bengal									
All	0.2320	878	0.2316	529.8012	527.2098	528.3812	0.9660	0.0215	0.1524
	0.0051	158	0.0047	170.4682	164.0081	162.7237	0.0026	0.0033	0.0051
Rural	0.1812	551	0.1811	337.3440	337.1524	337.1660	0.9666	0.0197	0.0546
	0.0048	138	0.0042	92.0419	91.5658	91.6090	0.0025	0.0028	0.0032
Urban	0.2609	327	0.2609	460.0384	325.2216	325.3178	0.9998	0.0255	0.9930
	0.0077	142	0.0007	85.7729	8.1272	0.0007	0.0002	0.0004	0.0012

Table S8. Replication results for two-stage procedure on 64th NSS Data ($\alpha = 5\%, \omega = 0.025, k \leq 4$)

Region	$\hat{G}_H$ / $se(\hat{G}_H)$	H / t	$\overline{\hat{G}}_{\tilde{Q}}$ / $se(\hat{G}_{\tilde{Q}})$	$\overline{Q^*}$ / $se(Q^*)$	$\overline{Q}$ / $se(Q)$	$\overline{\tilde{Q}}$ / $se(\tilde{Q})$	p / s_p	$\overline{w}_{\tilde{Q}}$ / $se(w_{\tilde{Q}})$	$p(w_{\tilde{Q}} > \omega)$ / $se(p(w_{\tilde{Q}} > \omega))$
Utter Pradesh									
All	0.2163	1262	0.2159	698.7892	695.6476	709.7158	0.9838	0.0213	0.1336
	0.0042	302	0.0040	254.9395	246.8035	234.0947	0.0018	0.0032	0.0048
Rural	0.1997	903	0.1993	476.1066	476.1066	475.9842	0.9838	0.0214	0.1758
	0.0041	197	0.0041	157.4488	157.4488	158.1605	0.0018	0.0035	0.0054
Urban	0.2229	359	0.2229	850.5696	358.9922	358.9976	1.0000	0.0362	1.0000
	0.0092	177	0.0002	212.7411	0.5515	0.1697	0.0000	0.0001	0.0000
West Bengal									
All	0.2320	878	0.2316	730.4786	686.8248	686.7118	0.9914	0.0225	0.2094
	0.0051	186	0.0033	232.2337	167.1106	167.6879	0.0013	0.0029	0.0058
Rural	0.1812	551	0.1809	452.0266	432.7092	432.7040	0.9928	0.0208	0.0688
	0.0048	163	0.0028	116.9002	87.0609	87.1087	0.0012	0.0024	0.0036
Urban	0.2609	327	0.2609	628.6992	327.0000	327.0000	1.0000	0.0303	1.0000
	0.0077	162	0.0000	104.9581	0.0000	0.0000	0.0000	0.0000	0.0000

Table S9. Simulation results for purely sequential procedure ($\alpha = 0.1$, $\omega = 0.02$, $m' = 1$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	273.3894 (41.2829)	0.9020 (0.0042)	0.0155 (0.0012)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	441.5382 (49.2216)	0.8944 (0.0043)	0.0184 (0.0003)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	403.3110 (25.5081)	0.8976 (0.0043)	0.0182 (0.0002)	0.0000 (0.0000)

2. Replication Using a Pseudo Population

First, a pseudo population is created considering the structure of the survey described in the main paper. This population is subdivided into three strata ($S = 3$, i.e. $s = 1, 2, 3$) with each stratum having 600 clusters ($H_s = 600$ for $s = 1, 2, 3$). The clusters are further divided into households. In all, the total number of households M is 900 000, and each sth stratum having $M_s = 300\,000$ number of households. Each household is randomly assigned a household size ν_{sc_sh}. To create three different synthetic populations, we allow the per capita monthly expenditure x_{sc_sh} for the households to follow a chosen theoretical distribution. The distributions chosen were Pareto(scale=20000, shape=5), Lognormal(mean = 2.185, sd = 0.562) and Gamma(shape = 2.649, rate = 0.84). The choice of the parameter values of the distributions are same as [3], [1], and [2].

For each of the datasets, the upper bound of the confidence interval width ω and the confidence level $1 - \alpha$ are pre-specified. The pilot sample size for each sth stratum is then computed by using the pilot sample size formula. For example, for $\omega = 0.02$ and $\alpha = 0.05$, we start with a pilot sample size of

$$t_s = \min\left\{600, \max\left\{2, \left\lceil \frac{2a_s z_{0.05/2}}{0.02} \right\rceil \right\}\right\} = 66$$

for $s = 1, 2, 3$. Thus, the initial number of clusters to be taken from the population is $t = \sum_{s=1}^{3} t_s = 198$. Now, a simple random sample of t_s clusters is drawn from stratum s. Then for each selected cluster, a simple random sample of $k = 4$ households is drawn. The per capita monthly expenditure x_{sc_sh} of the hth household belonging to the c_sth cluster from the s^{th} stratum is recorded and weighted with $w_{sc_sh} = M_{sc_s} H_s \nu_{sc_sh}/(kt_s)$. The estimator of ξ^2, V_t^2 is computed.

Now we apply the purely sequential procedure and the two-stage procedure for $\alpha \in \{0.05, 0.10\}$, $\omega \in \{0.02, 0.025\}$ and $m' \in \{1, 10, 20\}$ with replication size being 5000.

Tables S9 through S20 show the results of the Monte Carlo simulations on the three pseudo populations using the purely sequential procedure. The second column of the tables gives the theoretical Gini index (G) given the distribution in the first column. In the third column, $\bar{N}$ and s_N indicate the estimate of the average optimal sample size and its standard error respectively. The coverage probability (p) of the 5000 confidence intervals is in column 4 with its corresponding standard error being s_p. The average length of the confidence intervals ($\bar{w}_N$) is in column 5 and s_{w_N} is its standard error. The last column denotes the proportion of confidence intervals that were wider than the specified upper bound ω ($i.e.\ \Pr(w_N > \omega)$).

Results in Tables S9 to S20 show that the average widths were all less than the pre-specified value of ω. Since, $\Pr(w_N > \omega) = 0$ for all the results, the width of the confidence intervals will be less than ω. This was also indicated while replicating the procedure using NSS data. The coverage probabilities obtained were approximately equal to $100(1 - \alpha)\%$.

The results in Tables S21–S24 show the properties of the Monte Carlo simulations on the three pseudo populations using the two-stage methodology. The second column of the tables gives the theoretical Gini index (G) given the distribution in the first column. The average final sample size

Table S10. Simulation results for purely sequential procedure ($\alpha = 0.05$, $\omega = 0.02$, $m' = 1$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	354.2580 (55.5185)	0.9484 (0.0031)	0.0163 (0.0010)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	602.5182 (61.3798)	0.9448 (0.0032)	0.0188 (0.0002)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	546.0798 (32.2857)	0.9514 (0.0030)	0.0186 (0.0002)	0.0000 (0.0000)

Table S11. Simulation results for purely sequential procedure ($\alpha = 0.1$, $\omega = 0.025$, $m' = 1$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	199.3122 (29.6017)	0.8952 (0.0043)	0.0181 (0.0019)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	300.9054 (36.0870)	0.8942 (0.0043)	0.0223 (0.0006)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	278.0940 (18.6691)	0.8940 (0.0044)	0.0219 (0.0004)	0.0000 (0.0000)

Table S12. Simulation results for purely sequential procedure ($\alpha = 0.05$, $\omega = 0.025$, $m' = 1$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	253.7610 (38.1767)	0.9486 (0.0031)	0.0191 (0.0016)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	404.7870 (46.0777)	0.9408 (0.0033)	0.0229 (0.0005)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	370.9584 (23.9455)	0.9484 (0.0031)	0.0226 (0.0003)	0.0000 (0.0000)

Table S13. Simulation results for purely sequential procedure ($\alpha = 0.1$, $\omega = 0.02$, $m' = 10$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale = 20000, shape = 5)	0.1111	288.3120 (43.1982)	0.9020 (0.0042)	0.0152 (0.0012)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	456.3240 (49.8155)	0.8976 (0.0043)	0.0182 (0.0004)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	417.1740 (26.9280)	0.8950 (0.0043)	0.0179 (0.0003)	0.0000 (0.0000)

Table S14. Simulation results for purely sequential procedure ($\alpha = 0.05$, $\omega = 0.02$, $m' = 10$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale = 20000, shape = 5)	0.1111	368.7300 (55.9612)	0.9466 (0.0032)	0.0160 (0.0010)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	615.7560 (61.8045)	0.9450 (0.0032)	0.0186 (0.0003)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	558.6360 (33.2710)	0.9516 (0.0030)	0.0184 (0.0002)	0.0000 (0.0000)

Table S15. Simulation results for purely sequential procedure ($\alpha = 0.1$, $\omega = 0.025$, $m' = 10$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale = 20000, shape = 5)	0.1111	213.9060 (30.8347)	0.8996 (0.0043)	0.0175 (0.0018)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	315.6540 (37.4517)	0.8984 (0.0043)	0.0218 (0.0007)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	291.9000 (20.6367)	0.8924 (0.0044)	0.0213 (0.0006)	0.0000 (0.0000)

Table S16. Simulation results for purely sequential procedure ($\alpha = 0.05$, $\omega = 0.025$, $m' = 10$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale = 20000, shape = 5)	0.1111	268.6380 (39.6310)	0.9488 (0.0031)	0.0187 (0.0016)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	418.0260 (47.0880)	0.9400 (0.0034)	0.0226 (0.0005)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	383.4900 (25.1437)	0.9478 (0.0031)	0.0222 (0.0004)	0.0000 (0.0000)

Table S17. Simulation results for purely sequential procedure ($\alpha = 0.1$, $\omega = 0.02$, $m' = 20$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	305.1360 (45.3104)	0.8984 (0.0043)	0.0148 (0.0013)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	471.8640 (52.4671)	0.8980 (0.0043)	0.0179 (0.0005)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	432.0480 (31.5691)	0.8954 (0.0043)	0.0176 0.0004	0.0000 (0.0000)

Table S18. Simulation results for purely sequential procedure ($\alpha = 0.05$, $\omega = 0.02$, $m' = 20$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	384.3720 (59.4090)	0.9484 (0.0031)	0.0157 (0.0011)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	631.2240 (63.9250)	0.9442 (0.0032)	0.0184 (0.0004)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	573.5520 (36.2823)	0.9516 (0.0030)	0.0182 (0.0003)	0.0000 (0.0000)

Table S19. Simulation results for purely sequential procedure ($\alpha = 0.1$, $\omega = 0.025$, $m' = 20$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	226.2000 (37.8894)	0.8984 (0.0043)	0.0171 (0.0017)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	331.5240 (39.7815)	0.9010 (0.0042)	0.0213 (0.0009)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	308.9280 (20.8713)	0.8940 (0.0044)	0.0208 (0.0008)	0.0000 (0.0000)

Table S20. Simulation results for purely sequential procedure ($\alpha = 0.05$, $\omega = 0.025$, $m' = 20$)

Distribution	G	$\bar{N}$ (s_N)	p (s_p)	$\bar{w}_N$ (s_{w_N})	$\Pr(w_N > \omega)$ $se(\Pr(w_N > \omega))$
Pareto (scale=20000, shape=5)	0.1111	285.5040 (42.7734)	0.9480 (0.0031)	0.0182 (0.0017)	0.0000 (0.0000)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	433.5600 (49.7480)	0.9412 (0.0033)	0.0222 (0.0007)	0.0000 (0.0000)
Gamma (shape = 2.649, rate = 0.84)	0.3308	398.8440 (27.4951)	0.9530 (0.0030)	0.0218 (0.0006)	0.0000 (0.0000)

Table S21. Simulation results for two-stage procedure ($\alpha = 0.1$, $\omega = 0.02$)

Distribution	G	$\bar{Q}$ (s_Q)	p (s_p)	$\bar{w}_Q$ (s_{w_Q})	$\Pr(w_Q > \omega)$ $se(\Pr(w_Q > \omega))$
Pareto (scale = 20000, shape = 5)	0.1111	342.0360 (86.7413)	0.8996 (0.0043)	0.0141 (0.0017)	0.0068 (0.0012)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	546.3234 (97.0866)	0.8968 (0.0043)	0.0167 (0.0013)	0.0140 (0.0017)
Gamma (shape = 2.649, rate = 0.84)	0.3308	499.1904 (47.0225)	0.9034 (0.0042)	0.0164 (0.0007)	0.0000 (0.0000)

after 5000 replications is denoted as $\bar{Q}$ along with the corresponding standard error s_Q. The coverage probability (p) of the 5000 confidence intervals is in column 4 with its standard error being s_p. The average length of the confidence intervals ($\bar{w}_Q$) is in column 5 and s_{w_Q} is its standard error. The last column denotes the proportion of confidence intervals that were wider than the specified upper bound ω.

Results in Tables S21 through S24 show that the width of the confidence intervals may not be less than ω. This was also indicated while replicating the procedure using NSS data. However, the coverage probabilities obtained were approximately equal to $100(1-\alpha)\%$. The sample estimates for the Gini indices under the different simulation scenarios were tested for normality using Kolmogorv-Smirnov, Shapiro-Wilk, Anderson-Darling and Jarque-Bera Normailty tests and it was found that the normality assumption was not rejected.

References

1. Chattopadhyay, Bhargab and Shyamal Krishna De. 2016. Estimation of Gini index within pre-specified error bound. *Econometrics 4*(3), 30. doi:10.3390/econometrics4030030.
2. De, Shyamal K. and Bhargab Chattopadhyay. 2017. Minimum risk point estimation of Gini index. *Sankhya B 79*(2), 247–277. doi:10.1007/s13571-017-0140-3.

Table S22. Simulation results for two-stage procedure ($\alpha = 0.05$, $\omega = 0.02$)

Distribution	G	$\bar{Q}$ (s_Q)	p (s_p)	$\bar{w}_Q$ (s_{w_Q})	$\Pr(w_Q > \omega)$ $se(\Pr(w_Q > \omega))$
Pareto (scale = 20000, shape = 5)	0.1111	447.1206 (111.9348)	0.9474 (0.0032)	0.0148 (0.0017)	0.0112 (0.0056)
Lognormal (mean = 2.185, sd = 0.562)	0.3089	737.7720 (126.1189)	0.9432 (0.0033)	0.0172 (0.0012)	0.0190 (0.0019)
Gamma (shape = 2.649, rate = 0.84)	0.3308	670.0152 (60.5521)	0.9512 (0.0030)	0.0169 (0.0007)	0.0002 (0.0020)

Table S23. Simulation results for two-stage procedure ($\alpha = 0.1$, $\omega = 0.025$)

Distribution	G	$\bar{Q}$ (s_Q)	p (s_p)	$\bar{w}_Q$ (s_{w_Q})	$\Pr(w_Q > \omega)$ $se(\Pr(w_Q > \omega))$
Pareto	0.1111	246.1068	0.8980	0.0165	0.0052
(scale = 20000, shape = 5)		(63.1499)	(0.0043)	(0.0021)	(0.0010)
Lognormal	0.3089	375.8364	0.8968	0.0201	0.0098
(mean = 2.185, sd = 0.562)		(69.7613)	(0.0043)	(0.0016)	(0.0014)
Gamma	0.3308	345.9612	0.8990	0.0197	0.0000
(shape = 2.649, rate = 0.84)		(33.8431)	(0.0043)	(0.0008)	(0.0000)

Table S24. Simulation results for two-stage procedure ($\alpha = 5\%$, $\omega = 0.025$)

Distribution	G	$\bar{Q}$ (s_Q)	p (s_p)	$\bar{w}_Q$ (s_{w_Q})	$\Pr(w_Q > \omega)$ $se(\Pr(w_Q > \omega))$
Pareto	0.1111	316.3188	0.9476	0.0174	0.0068
(scale = 20000, shape = 5)		(79.7332)	(0.0032)	(0.0021)	(0.0012)
Lognormal	0.3089	502.0482	0.9454	0.0208	0.0120
(mean = 2.185, sd = 0.562)		(90.2360)	(0.0032)	(0.0016)	(0.0015)
Gamma	0.3308	459.2406	0.9512	0.0204	0.0000
(shape = 2.649, rate = 0.84)		(43.7022)	(0.0030)	(0.0009)	(0.0000)

3. Ransom, Michael R and Jan S Cramer. 1983. Income distribution functions with disturbances. *European Economic Review 22*(3), 363–372. doi:10.1016/0014-2921(83)90050-8.