

BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-86/17

A Study of Whether Census
Adjustment is Worthwhile

by

Beverley Causey
Statistical Research Division
Bureau of the Census

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Kirk M. Wolter

Report completed: July 9, 1986

Report issued: July 9, 1986

A Study of Whether Census Adjustment is Worthwhile

1. INTRODUCTION

Suppose that we have a set of N disjoint blocks, indexed $i=1, \dots, N$. One might think of these blocks as being the 50 states plus DC, but they might also correspond to smaller areas. For block i we let t_i denote true population, and y_i denote population as officially obtained from the Census. Let $T = \sum t_i$ and $Y = \sum y_i$. We assume knowledge of y_i (and thus Y) and T , but not of t_i . Presumably the knowledge of T would come from an external source such as demographic methods.

Consider the ratio adjustment across the board:

$$a_i = y_i T/Y, \quad (1)$$

which removes the effects of overall (as opposed to differential) relative overcount or undercount. We want to examine whether a_i is better than y_i : "better" in the sense that a_i is closer to t_i than is y_i . We will consider four criteria for closeness. These criteria are as follows.

Let \underline{a} denote the vector (a_1, \dots, a_N) , with \underline{y} and \underline{t} likewise; and let \underline{x} denote the (arbitrary) vector (x_1, \dots, x_N) . With \underline{y} and \underline{t} fixed, let

$$f_1(\underline{x}) = \sum_{i=1}^N (x_i - t_i)^2 \quad (2a)$$

$$f_2(\underline{x}) = \sum (x_i - t_i)^2 / y_i \quad (2b)$$

$$f_3(\underline{x}) = \sum (x_i - t_i)^2 / t_i \quad (2c)$$

$$f_4(\underline{x}) = \sum |x_i - t_i|. \quad (2d)$$

These are our four criteria, or loss functions. For $k = 1, 2, 3, 4$, we want to compare $f_k(\underline{a})$ and $f_k(\underline{y})$. We will examine the extent to which it is the case that

$$f_k(\underline{a}) < f_k(\underline{y}). \quad (3)$$

We will consider two approaches: in Section 2 we deal with mathematical analysis, and in Section 3 we deal with empirical results which, possibly, are indicative of what happens in practice. Section 4 is a summary.

2. MATHEMATICAL ANALYSIS

* There is no convenient characterization as to when the inequality in (3) holds for $k=4$. For $k=2$ it always holds: that is,

$$f_2(\underline{a}) < f_2(\underline{y}) \quad (4)$$

always.

We may easily demonstrate (4) by letting $\underline{x} = c\underline{y}$ for varying scalar c and establishing that $f_2(\underline{x})$ attains a minimum uniquely for $c = T/Y$.

For $k=1$, let

$$R = T/Y \text{ and } B = (\sum t_i y_i) / (\sum y_i^2). \quad (5)$$

Then, we have

$$f_1(\underline{a}) < f_1(\underline{y}) \text{ if and only if } |R-B| < |1-B|. \quad (6)$$

Proof of (6) may be based on, again, letting $\underline{x} = c\underline{y}$ for varying scalar c and viewing f_1 as a quadratic function of c which attains a minimum at $c=B$.

Here is a simple example where the inequalities in (6) do not hold, for 2 states:

$$t_1=2, t_2=8/3 = 2 \frac{2}{3}, y_1=1, y_2=3, B=1, R=7/6. \quad (7)$$

Thus (3) does not always hold for $k=1$. It does not hold when $B=1$ and $R \neq 1$. To have $B=1$ we must have a balancing between some y_i overstating t_i and some y_i understating it. But when such balancing occurs, R will typically also be close to 1. In practice, one might guess that B will not be close to 1, however, largely because there has been a prevailing pattern of y_i understating t_i .

For $k=3$ we set

$$B^* = Y / (\sum y_i^2 / t_i) \quad (8)$$

and obtain, in analogy with (6), the characterization

$$f_3(\underline{a}) < f_3(\underline{y}) \text{ if and only if } |R-B^*| < |1-B^*|. \quad (9)$$

3. EMPIRICAL RESULTS

We now examine empirically the extent to which the 4 inequalities in (3) are satisfied.

We consider results from the 1980 Decennial Census (y_i) and the Post-Enumeration Survey (PES) for that census. This followup coverage survey is carried out for a nationwide sample of persons, and produces estimators of the true but unknown t_i , with associated sampling variances which too have been estimated. Our approach, essentially, is (a) to act like these estimators are the true t_i , and (b) in the light of the sampling error that is indicated by our sampling variance, look at the statistical significance of these findings.

Thus, let \hat{t}_i denote the estimator of t_i . We first form $\hat{T} = \{\hat{t}_i\}$ and act as though \hat{T} is the true T that would be obtained from external demographic methods. Thus we form a_i based on ratio adjustment in (1). Then, to compute the loss functions f_k , we substitute \hat{t}_i for t_i . Based on these substitutions, we set

$$\hat{E}_k = \hat{f}_k(\underline{a})/\hat{f}_k(\underline{y}) \quad (10)$$

as indication of how much we gain, relatively, by using a_i instead of y_i as a proxy for t_i .

Under different assumptions, 12 sets of PEP estimators (with accompanying variances) were formed. Thus we obtain 12 sets of quantities \hat{E}_k :

set	\hat{E}_1	\hat{E}_2	\hat{E}_3	\hat{E}_4
2-8	.454	.611	.619	.732
10-8	.820	.928	.934	.961
2-9	.374	.499	.506	.646
3-8	.507	.677	.686	.804
3-9	.407	.548	.556	.706
5-8	.293	.406	.412	.589
5-9	.253	.345	.350	.551
3-20	.343	.470	.475	.634
2-20	.319	.429	.435	.586
14-8	.686	.046	.639	.760
14-9	.986	.865	.857	.882
14-20	1.031	.983	.979	.988

Thus only the last set, 14-20, for f_1 provides any indication that a_i is not an improvement over y_i .

To assess the statistical significance of these results, we considered differences

$$\hat{D}_k = \hat{f}_k(\underline{y}) - \hat{f}_k(\underline{a}). \quad (11)$$

Corresponding to these differences there are, of course, underlying parameters D_k . We tested the null hypothesis $D_k = 0$ (no difference between \underline{y} and \underline{a}) against the alternative $D_k > 0$ (\underline{a} is an improvement). These hypotheses correspond, of course, to $E_k = 1$ and $E_k < 1$ respectively. For $k=2$ we already know $D_2 > 0$ as in (4); thus we consider the test only for $k=1, 3$ and 4.

We may form an approximation to $\text{Var}(\hat{D}_k)$ based on a 1st-order Taylor linearization of \hat{f}_k along with the aforementioned estimators of the quantities $\text{Var}(\hat{t}_1)$. (In the case of $\text{Var}(\hat{D}_4)$ involving absolute values, this variance approximation is not completely conventional; but results for \hat{D}_4 , as we shall

discuss, correspond to those for \hat{D}_1 and \hat{D}_3 .) We divide \hat{D}_k by its estimated standard deviation and compare the resulting test t-statistic against $N(0,1)$ to test the null hypothesis. For 9 of the 12 sets of PEP estimators (all but sets 10-8, 14-9, and 14-20), the (positive) difference \hat{D}_k was statistically significant at the 5% level. For set 14-20 and D_3 , also, the difference was significant.

These, then, are the extent of the empirical findings to support the claim that \underline{a} should be used in preference to \underline{y} . For all four of the criterion f_k in (2) this claim appears to hold up.

4. SUMMARY

We have shown that under a variety of standard loss functions, the simple ratio adjustment (1) moves the data closer to the truth. This result does not appear to be sensitive to the choice of loss function, thus confirming observations made in the CNSTAT final report.

The key assumption in the considered ratio adjustment is that the true total, T , is known and can be exploited in the ratio adjustment. Future reports will examine the extent to which the present findings hold true when the ratio adjustment is based upon an estimator \hat{T} of T with known distributional properties.

1980 CENSUS COUNTS

ST_60	ST_NAME	CEN_CT
11	MAINE	1124660
12	NEW HAMPSHIRE	920610
13	VERMONT	511456
14	MASSACHUSETTS	5737037
15	RHODE ISLAND	947154
16	CONNECTICUT	3107576
21	NEW YORK	17558072
22	NEW JERSEY	7364823
23	PENNSYLVANIA	11863895
31	OHIO	10797630
32	INDIANA	5490224
33	ILLINOIS	11426518
34	MICHIGAN	9262078
35	WISCONSIN	4705767
41	MINNESOTA	4075970
42	IOWA	2913808
43	MISSOURI	4916686
44	NORTH DAKOTA	652717
45	SOUTH DAKOTA	690768
46	NEBRASKA	1569825
47	KANSAS	2363679
51	DELAWARE	594338
52	MARYLAND	4216975
53	DISTRICT OF COL.	638333
54	VIRGINIA	5346818
55	WEST VIRGINIA	1949644
56	NORTH CAROLINA	5881766
57	SOUTH CAROLINA	3121820
58	GEORGIA	5463105
59	FLORIDA	9746324
61	KENTUCKY	3660777
62	TENNESSEE	4591120
63	ALABAMA	3893888
64	MISSISSIPPI	2520638
71	ARKANSAS	2286435
72	LOUISIANA	4205900
73	OKLAHOMA	3025290
74	TEXAS	14229191
81	MONTANA	786690
82	IDAHO	943935
83	WYOMING	469557
84	COLORADO	2889964
85	NEW MEXICO	1302894
86	ARIZONA	2718215
87	UTAH	1461037
88	NEVADA	800493
91	WASHINGTON	4132156
92	OREGON	2633105
93	CALIFORNIA	23667902
94	ALASKA	401851
95	HAWAII	964691

From PC80-1-1A, U. S. Summary, Table 8, p. 43.

PEP ESTIMATES AND THEIR STANDARD ERRORS

ST_NAME	P_14_8	SE_14_8	P_14_9	SE_14_9	P_14_20	SE_14_20
MAINE	1133442	11779	1133819	11686	1136592	11526
NEW HAMPSHIRE	900058	9212	901668	9173	905632	9593
VERMONT	504598	2607	503889	2825	507410	2947
MASSACHUSETTS	5610009	33607	5659372	31177	5675444	31965
RHODE ISLAND	940363	5830	939383	5611	943048	5065
CONNECTICUT	3046416	35128	3057298	34143	3071513	27244
NEW YORK	17383935	79571	17468341	82674	17510441	83021
NEW JERSEY	7349244	49201	7348474	49534	7376830	51949
PENNSYLVANIA	11733989	55687	11763321	56751	11789261	56369
OHIO	10750393	57420	10761895	58527	10781745	56147
INDIANA	5385405	38088	5389839	37434	5397001	36246
ILLINOIS	11327722	55281	11407278	58554	11423435	57569
MICHIGAN	9217382	41963	9250276	40324	9270978	41627
WISCONSIN	4739499	21763	4739499	21763	4740968	22426
MINNESOTA	4083163	22629	4098277	23621	4102679	21977
IOWA	2875557	31942	2875557	31942	2876769	33202
MISSOURI	4911352	32186	4916254	31629	4928215	30591
NORTH DAKOTA	647731	2390	649538	2308	649691	2323
SOUTH DAKOTA	685668	2748	686531	2461	687216	2541
NEBRASKA	1557977	7354	1561192	7194	1564690	6476
KANSAS	2343775	16682	2347907	16171	2356056	14618
DELAWARE	583846	6031	585168	5631	587416	5421
MARYLAND	4254937	42689	4265835	41272	4276896	40938
DISTRICT OF COL.	647834	10791	661198	10317	669127	9577
VIRGINIA	5300229	43167	5323099	42628	5334276	43337
WEST VIRGINIA	1912683	27574	1912188	27343	1918131	26380
NORTH CAROLINA	5864556	65152	5877425	60004	5899525	57015
SOUTH CAROLINA	3234299	76660	3246235	76095	3281862	74909
GEORGIA	5346937	34325	5370092	34716	5388033	39923
FLORIDA	9647620	71854	9702921	74012	9746240	68120
KENTUCKY	3567481	28063	3567659	30468	3587015	28216
TENNESSEE	4410031	55286	4434256	55235	4431468	55573
ALABAMA	3807843	32152	3820235	30687	3843447	32968
MISSISSIPPI	2498395	28288	2512079	25706	2516300	27645
ARKANSAS	2234157	31747	2239942	30547	2254683	30982
LOUISIANA	4185817	90828	4196851	81061	4228070	81332
OKLAHOMA	2946071	20237	2946844	19648	2958441	20276
TEXAS	13778895	101087	13890195	97672	13958898	99513
MONTANA	785850	6550	788038	6310	791057	7056
IDAHO	946492	6060	947605	5858	950557	5058
WYOMING	477834	4666	479803	5210	479958	5176
COLORADO	2842810	22065	2861295	24070	2869023	24890
NEW MEXICO	1287808	11085	1291059	11246	1295591	11518
ARIZONA	2716403	35819	2738359	33943	2752653	29879
UTAH	1459520	9701	1469576	9265	1470772	9727
NEVADA	799035	10117	816468	9694	821234	9543
WASHINGTON	4112177	29036	4110050	29282	4132475	25471
OREGON	2602181	15713	2614882	14213	2624861	12502
CALIFORNIA	23747738	100346	23935428	96609	24018862	93470
ALASKA	406537	4784	408387	4871	409122	4821
HAWAII	955198	8986	956720	9017	962707	8191

PEP ESTIMATES AND THEIR STANDARD ERRORS

ST_NAME	P_5_9	SE_5_9	P_3_20	SE_3_20	P_2_20	SE_2_20
MAINE	1147827	9672	1151422	10263	1152058	10127
NEW HAMPSHIRE	932986	12344	912601	11325	912068	11080
VERMONT	510112	3876	509164	2892	509899	3208
MASSACHUSETTS	5760173	50760	5744186	35216	5744774	34312
RHODE ISLAND	958391	8389	959387	8473	958985	8478
CONNECTICUT	3136759	36073	3103238	30933	3118575	33470
NEW YORK	17958646	120366	18001262	110647	18031802	108927
NEW JERSEY	7479318	64765	7497216	68247	7509603	68687
PENNSYLVANIA	11877050	78827	11902607	63505	11943069	66392
OHIO	10923983	53363	10946193	65204	10948088	65093
INDIANA	5625813	50314	5467317	42094	5467819	42557
ILLINOIS	11639057	84989	11766850	70260	11758411	67853
MICHIGAN	9407742	39569	9391797	54750	9410438	56899
WISCONSIN	4721516	16949	4786560	26549	4785651	27184
MINNESOTA	4134647	19323	4138130	22531	4142497	24089
IOWA	2933860	24101	2894572	32219	2897219	33413
MISSOURI	4948854	34852	4972828	35803	4981497	34831
NORTH DAKOTA	657974	3704	654538	2050	654649	2013
SOUTH DAKOTA	705992	7697	691991	3475	693427	3783
NEBRASKA	1593988	15170	1577211	8702	1578417	7740
KANSAS	2411717	20325	2388642	21119	2393672	22649
DELAWARE	602116	6859	594523	5938	595915	5301
MARYLAND	4349139	33640	4341398	55187	4350610	55976
DISTRICT OF COL.	655310	14006	684647	10264	687044	9864
VIRGINIA	5463288	39469	5387357	36819	5419582	48845
WEST VIRGINIA	1929721	24188	1943662	24935	1942108	24187
NORTH CAROLINA	6019022	72405	5987293	58600	5985507	61982
SOUTH CAROLINA	3265538	93135	3366650	90400	3362934	89108
GEORGIA	5561556	43948	5482471	45529	5503578	44857
FLORIDA	10226444	160251	9985520	78117	10006779	74771
KENTUCKY	3667009	39447	3623860	34627	3629793	34145
TENNESSEE	4511838	68109	4484931	57244	4485744	56168
ALABAMA	3892997	32017	3917646	40584	3917039	39660
MISSISSIPPI	2611013	43295	2563415	30642	2563906	31345
ARKANSAS	2324880	36342	2283101	27728	2284426	27433
LOUISIANA	4296275	88040	4348857	85508	4344365	84488
OKLAHOMA	3050086	41578	3027835	29481	3030504	29654
TEXAS	14759062	158065	14459291	126928	14499190	127876
MONTANA	809072	8134	803893	9269	804796	8747
IDAHO	989582	10876	960694	5709	962329	6418
WYOMING	493962	6840	489130	6900	489794	6683
COLORADO	3018413	35212	2928879	28342	2939314	30726
NEW MEXICO	1355898	19082	1340313	13047	1341986	13527
ARIZONA	2880359	34317	2811024	39522	2802321	35798
UTAH	1500886	10484	1479094	11227	1479989	11143
NEVADA	859325	15313	845656	10289	848574	10570
WASHINGTON	4281463	55524	4216774	36531	4222592	34746
OREGON	2707216	25585	2668734	17082	2671585	17659
CALIFORNIA	24568211	133003	24687815	126695	24760391	127307
ALASKA	435264	11910	415678	5071	416155	5063
HAWAII	986124	9925	980799	11648	982873	10894

PEP ESTIMATES AND THEIR STANDARD ERRORS

ST_NAME	P_2_8	SE_2_8	P_10_8	SE_10_8	P_2_9	SE_2_9
MAINE	1148857	10381	1138585	10386	1149257	10427
NEW HAMPSHIRE	906427	10716	913701	13818	908053	10716
VERMONT	507065	2840	503948	2558	506353	3120
MASSACHUSETTS	5677995	36522	5665354	50929	5728378	33614
RHODE ISLAND	956158	8782	939519	7170	955215	8723
CONNECTICUT	3092482	41701	3066917	32045	3103845	40463
NEW YORK	17896310	106844	17649139	116359	17986995	109157
NEW JERSEY	7480686	66647	7376197	50172	7480049	66737
PENNSYLVANIA	11886284	66475	11728624	71873	11916546	66299
OHIO	10915915	64874	10803257	48353	10928111	66213
INDIANA	5456021	44338	5573542	56715	5460497	44011
ILLINOIS	11658112	67217	11426594	74179	11741166	68136
MICHIGAN	9354474	54896	9252240	36172	9389139	54405
WISCONSIN	4784191	26437	4703058	18116	4784191	26437
MINNESOTA	4122914	25671	4092346	17679	4138069	25781
IOWA	2896006	32207	2903087	22340	2896006	32207
MISSOURI	4964266	36444	4890199	31127	4969316	35812
NORTH DAKOTA	652652	2056	654444	3323	654487	2020
SOUTH DAKOTA	691857	3946	696137	5994	692730	3729
NEBRASKA	1571571	8312	1576922	13337	1574864	8311
KANSAS	2381114	23562	2384627	23737	2385339	23300
DELAWARE	592297	5796	595941	7303	593623	5488
MARYLAND	4328315	57282	4291222	38343	4339304	56077
DISTRICT OF COL.	665112	11290	633892	11858	678912	10457
VIRGINIA	5384287	49025	5392551	38061	5407359	47894
WEST VIRGINIA	1936644	24993	1908116	24119	1936108	24764
NORTH CAROLINA	5949657	68494	5951662	71751	5962736	64359
SOUTH CAROLINA	3314170	89094	3191081	82711	3326403	89255
GEORGIA	5461848	40043	5413506	42547	5485548	39993
FLORIDA	9905683	78574	9876934	142058	9962705	80150
KENTUCKY	3610017	33671	3633443	46685	3610074	36183
TENNESSEE	4463364	56031	4444731	65572	4488596	55749
ALABAMA	3880007	36012	3844577	33718	3892815	35031
MISSISSIPPI	2545731	32163	2545183	37082	2559566	29542
ARKANSAS	2263733	28114	2285270	41867	2269627	26818
LOUISIANA	4300078	94091	4215907	84803	4312276	82796
OKLAHOMA	3017843	29738	3003971	30937	3018609	29470
TEXAS	14309405	127810	14257990	138549	14427264	125196
MONTANA	799391	8091	803923	8015	801662	8006
IDAHO	958180	7354	971455	8145	959307	7141
WYOMING	487629	6273	480384	4684	489633	6706
COLORADO	2912128	27870	2935075	31971	2931335	29258
NEW MEXICO	1333681	12906	1333274	16547	1337235	13250
ARIZONA	2765268	40617	2790879	35949	2787619	39204
UTAH	1468643	11285	1466806	9573	1478797	10715
NEVADA	825059	10851	808559	8405	843642	10635
WASHINGTON	4201690	38090	4204009	41839	4199458	38472
OREGON	2646687	20366	2657919	23235	2661388	19288
CALIFORNIA	24476807	137350	23975957	98559	24673911	131605
ALASKA	413497	5018	420221	12231	415397	5066
HAWAII	975149	11522	974713	9114	976712	11697

PEP ESTIMATES AND THEIR STANDARD ERRORS

ST_NAME	P_3_8	SE_3_8	P_3_9	SE_3_9	P_5_8	SE_5_8
MAINE	1148216	10520	1148618	10571	1147413	9489
NEW HAMPSHIRE	906955	10954	908581	10960	931344	12401
VERMONT	506337	2544	505627	2828	510857	3781
MASSACHUSETTS	5677298	36982	5727773	34516	5708805	51516
RHODE ISLAND	956563	8742	955619	8681	959338	8362
CONNECTICUT	3076773	39515	3088168	38013	3124597	36053
NEW YORK	17866180	108180	17956517	110879	17870500	121754
NEW JERSEY	7468431	66195	7467772	66281	7480190	64326
PENNSYLVANIA	11846230	62738	11876188	63048	11847508	79294
OHIO	10914072	64921	10926238	66165	10911827	51266
INDIANA	5455524	43752	5459995	43420	5621245	50727
ILLINOIS	11665181	70695	11749424	71247	11553854	82670
MICHIGAN	9335928	53063	9370570	52449	9373506	41255
WISCONSIN	4785099	25802	4785099	25802	4721516	16949
MINNESOTA	4118554	24481	4133699	24046	4119626	17918
IOWA	2893358	30967	2893358	30967	2933860	24101
MISSOURI	4955711	37406	4960742	36845	4943710	35203
NORTH DAKOTA	652542	2094	654376	2054	656143	3687
SOUTH DAKOTA	690419	3632	691294	3415	705093	7696
NEBRASKA	1570371	9269	1573661	9298	1590660	15102
KANSAS	2376165	21878	2380331	21612	2407267	21016
DELAWARE	590896	6392	592224	6090	600709	7466
MARYLAND	4319123	56499	4330107	55276	4338196	34417
DISTRICT OF COL.	662779	11490	676546	10762	642265	13887
VIRGINIA	5352941	38036	5376010	36109	5439742	38168
WEST VIRGINIA	1938186	25743	1937652	25518	1930327	24370
NORTH CAROLINA	5951617	65091	5964746	61120	6005549	72485
SOUTH CAROLINA	3317900	90296	3330163	90492	3253625	93790
GEORGIA	5440800	39961	5464384	40696	5537553	44147
FLORIDA	9884595	81449	9941468	82707	10166367	152275
KENTUCKY	3604103	34242	3604170	36794	3667319	38668
TENNESSEE	4462549	57078	4487767	56819	4485907	68145
ALABAMA	3880560	36878	3893386	35922	3879983	33336
MISSISSIPPI	2545228	31466	2559065	28843	2596972	42867
ARKANSAS	2262397	28578	2268294	27199	2318834	36328
LOUISIANA	4304462	95032	4316711	83583	4286482	96208
OKLAHOMA	3015205	29947	3015969	29645	3049244	41975
TEXAS	14270589	127313	14387681	124183	14639509	159466
MONTANA	798487	8677	800740	8487	806774	7886
IDAHO	956553	6531	957678	6343	988388	10608
WYOMING	486972	6410	488969	6932	491902	6849
COLORADO	2901640	26509	2920873	27758	2998616	33625
NEW MEXICO	1332019	12487	1335557	12727	1352069	18452
ARIZONA	2773884	44374	2796236	42887	2856920	36018
UTAH	1467747	11271	1477903	10805	1490449	10671
NEVADA	822213	10815	840745	10392	840821	15075
WASHINGTON	4195887	39938	4193657	40305	4283789	55210
OREGON	2643859	19620	2658535	18761	2693344	25410
CALIFORNIA	24405570	137305	24601506	131962	24373924	132392
ALASKA	413028	5042	414924	5064	433234	11984
HAWAII	973143	12109	974701	12345	984593	10280