POST-STRATIFICATION APPROACHES IN THE CORPORATION STATISTICS OF INCOME PROGRAM

James M. Harte, Internal Revenue Service

This paper is an early interim report on the research being conducted to improve estimation procedures in the Corporate Statistics of Income (SOI) program. This report provides just a brief sketch of the background for the application issues we expect to experience in more detail at next year's meetings.

The motivation for the present work is quite simple. Budget cuts have increased concerns that the corporation return sample size is inadequate. Raking ratio estimation and other post-stratification techniques are among the procedures being considered to improve the efficiency of the corporate sample for the primary statistics of interest. [1]

BACKGROUND

Annual statistics have been available from corporate tax returns since the 1916 income tax year. Figure 1 lists the major uses and publications of these data. The corporation source book is the most detailed SOI report featuring complete income statement and balance sheet information classified by industry. The three levels of classes are: industry division (12 classes), major industry (58 classes), and minor industry (160 classes). Figure 2 provides a specimen page from the source book for minor industry 2096. The names of the industry division, major industry, and minor industry are given at the bottom of the Figure.

A pilot study of the corporate SOI data was conducted some years ago where estimation was based upon post-stratification by industry [2] (see Figure 3). The study indicated that post-stratification by major industry could achieve large reductions in variance for some items and little, if any, increase in variance for others. The comparison was with a scheme in which tax returns were stratified by the joint size of their income and assets (essentially the same way as is done at present).

The full scale study we are now conducting involves all 1979 corporation tax returns filed on Form 1120 or Form 1120-S which were fractionally sampled. These are the two return types filed by the overwhelming majority of corporations (over 98%). Excluded from the study are the other corporation tax returns tabulated in SOI namely Forms 1120F, 1120L, 1120M and 1120 DISC. Excluded as well are those Forms 1120 and 1120S cases selected with certainty because of high income or high assets (or because they were needed for special studies).

Figure 4 shows how we classify the returns into sample strata. [3] In effect, we assign each return in our study a numerical code of 1 through 9 based upon its size of net income or deficit. We also assign each return a code of 1 through 9 based upon the size of its assets. The larger of the two codes labels the stratum in which the return is sampled at a given rate. Figure 5 shows the resultant distribution of the population and the sample by stratum. In our post-stratification studies these sample classes are further divided into 58 industry groups, producing 58 x 9 = 522 post-strata.

POST-STRATIFICATION

Figure 6 compares and contrasts the present stratification scheme with the post-stratification classification scheme used in our study. Four steps are listed for stratification with three of the four having a parallel step in post-stratification. The missing step is step 2, sample selection, because post-stratification is an estimation method and does not involve the selection of the sample. Item 3 under post-stratification actually describes the method of estimation used in the pilot study by Westat [2]. The known counts for the major industries are from the revenue processing of the returns and are based on the Principal Business Activity (PBA) Code. Unfortunately, industry post-stratification in its simplest version is unwieldy and can actually do more harm than good because the sample sizes in the 522 post-strata can be very small. Grouping of the sample into large enough categories to avoid this problem, as was done by Westat, is a very difficult procedure to do well and contains many arbitrary elements; hence we rejected it in favor of a "raking" approach.

Figure 1

CORPORATION INCOME TAX RETURN DATA
MAJOR USES AND SOURCES

USES

Revenue estimation and tax policy by Treasury Department and the Congress.

Estimates needed to produce the National Income and Product Accounts by the Commerce Department.

Information for business and industry analysts and economists (in both the private and public sector).

SOURCES

Statistics of Income Corporation Income Tax Returns, Internal Revenue Service, Publication 16 (annual publication available from the Government Printing Office).

Corporation Source Book of Statistics of Income (unpublished tables by industry group including minor industry; available by special order on a reimbursable basis from the Statistics of Income Division, Internal Revenue Service, Washington, DC).

Figure 2

SAMPLE PAGE FROM CORPORATION STATISTICS OF INCOME SOURCE BOOK FOR TAX YEAR 1979

MINOR INDUSTRY 2096 *: RETURNS WITH AND WITHOUT NET INCOME	SIZE OF TOTAL ASSETS												
	TOTAL	ZERO ASSETS	1 UNDER 100	100 UNDER 250	250 UNDER 500	500 UNDER 1,000	1,000 UNDER 5,000	5,000 UNDER 10,000	10,000 UNDER 25,000	25,000 UNDER 50,000	50,000 UNDER 100,000	100,000 UNDER 250,000	250,000 OR MORE
1 NUMBER OF RETURNS	3535	8	*1262	*694	555	367	414	110	69	28	15	7	6
2 TOTAL ASSETS	13185240	-	*70286	*118204	209740	281923	896121	775946	1115010	1026448	1081546	1015390	6594627
3 CASH	543054	-	*7687	*13919	11533	7139	48667	77429	64191	38615	32433	97500	143740
4 NOTES AND ACCOUNTS RECEIVABLE	2524409	-	*12375	*48059	50482	89011	203610	134584	271087	187125	252038	117722	1158315
5 LESS: ALLOWANCE FOR BAD DEBTS	54568	-	-	-	-	*883	8516	3917	4200	5218	6669	2650	22516
6 INVENTORIES	3102470	-	*14305	*13326	*34204	61879	240384	189906	296926	258913	209867	162095	1620665
INVESTMENTS IN GOVT. OBLIGATIONS:													
7 UNITED STATES	59692	-	-	-	-	*7387	1875	-	950	3516	918	17218	27828
8 STATE AND LOCAL	46975	-	-	-	-	-	-	-	1356	13936	694	-	30989
9 OTHER CURRENT ASSETS	646301	-	*2054	*3338	*2895	18379	40388	19934	46735	69972	66749	44052	331807
10 LOANS TO STOCKHOLDERS	52110	-	-	*2573	*457	*8423	*3914	*5275	9080	3752	5884	12751	-
11 MORTGAGE AND REAL ESTATE LOANS	*7909	-	-	-	*2788	-	-	*108	25	210	-	4778	-
12 OTHER INVESTMENTS	1833516	-	-	*3192	*33748	*9608	19749	41577	55511	105483	146869	199082	1218698
13 DEPRECIABLE ASSETS	6695317	-	*41282	*79144	120067	203498	561832	457157	583739	497845	534191	344207	3272356
14 LESS: ACCUMULATED DEPRECIATION	2804322	-	*16461	*49930	51658	134981	264727	189310	254244	211105	226401	104179	1301327
15 DEPLETABLE ASSETS	520	-	-	-	-	-	-	-	141	-	-	380	-
16 LESS: ACCUMULATED DEPLETION	83	-	-	-	-	-	-	-	83	-	-	-	-
17 LAND	215318	-	-	*3395	*1846	*3806	28856	16953	26158	23096	25367	15757	70084
18 INTANGIBLE ASSETS (AMORTIZABLE)	98815	-	*5312	*11819	-	*1028	*12730	8515	2972	3906	25315	5492	21726
19 LESS: ACCUMULATED AMORTIZATION	35463	-	*354	*11031	-	*1001	*5033	3220	1586	1618	1894	1738	7988
20 OTHER ASSETS	253270	-	*3885	*401	*3380	*8631	12390	20955	16253	38021	16184	102922	30250
21 TOTAL LIABILITIES	13185240	-	*70286	*118204	209740	281923	896121	775946	1115010	1026448	1081546	1015390	6594627
22 ACCOUNTS PAYABLE	1618058	-	*16549	*11557	34626	78546	186290	105923	151737	124662	136248	101809	670112
23 MORT, NOTES, AND BONDS UNDER 1 YR.	1298141	-	*12621	*223	*22639	39779	145580	101553	222774	137335	198281	69923	347433
24 OTHER CURRENT LIABILITIES	1409380	-	*1054	*9505	*21485	11907	61867	50914	77132	67168	71333	116890	920125
25 LOANS FROM STOCKHOLDERS	130564	-	*10949	*1255	*3839	*11765	19737	*5254	18639	22422	36704	-	-
26 MORT, NOTES, BONDS, 1 YR OR MORE	2285468	-	*21933	*3067	*25900	*26224	150688	146322	152009	203539	216488	193521	1145776
27 OTHER LIABILITIES	301774	-	*2761	*-0	*1720	*10753	25196	1964	21569	14712	14553	31868	176677
28 CAPITAL STOCK	1060444	-	*21467	*17278	24555	32274	66463	47153	101773	96727	97651	39515	515589
29 PAID-IN OR CAPITAL SURPLUS	1080254	-	-	*549	-	*1550	34049	75127	31076	61110	76339	160721	639735
30 RETAINED EARNINGS, APPROPRIATED	296156	-	-	-	-	-	-	*21	23461	12639	91947	-	168087
31 RETAINED EARNINGS, UNAPPROPRIATED	3870218	-	*-17047	*80758	*80613	70151	243279	256702	321766	304698	144977	301266	2083054
32 LESS: COST OF TREASURY STOCK	165217	-	-	*5988	*5638	*1023	37028	*14989	6928	18565	2975	122	71960
33 TOTAL RECEIPTS	30340273	322894	*216492	*339367	539787	862336	2468157	2020373	2498109	2456715	2177457	1561393	14877194
34 BUSINESS RECEIPTS	29711748	317711	*215933	*337477	529872	849703	2416133	1980161	2457514	2419854	2112002	1487206	14588183
INTEREST ON GOVT. OBLIGATIONS:													
35 UNITED STATES	5486	-	-	-	-	*209	-	211	562	751	126	1608	2018
36 STATE AND LOCAL	6062	-	-	-	-	-	*38	-	79	982	23	-	4940
37 OTHER INTEREST	169533	1017	*518	*1145	*5920	*923	3119	7829	13405	10501	10883	12817	101456
38 RENTS	39390	227	-	-	*1259	*3247	5898	4507	3653	2311	4573	1329	12386
39 ROYALTIES	25705	191	-	-	-	-	*250	*77	273	598	236	371	23709
40 NET S-T CAP GAIN LESS NET L-T LOSS	*4885	-	-	-	-	-	*15	4	352	11	-	4049	456
41 NET L-T CAP GAIN LESS NET S-T LOSS	87360	26	-	*110	*1436	*4611	12330	*14433	2146	1774	13519	30252	6723
42 NET GAIN, NONCAPITAL ASSETS	18468	527	-	*291	*93	*1393	786	932	1209	2651	3564	5549	1474
43 DIVIDENDS, DOMESTIC CORPORATIONS	74542	503	-	*344	-	-	*1	*3000	3855	4421	14534	4589	43295
44 DIVIDENDS, FOREIGN CORPORATIONS	41333	-	-	-	-	-	-	-	30	103	4158	269	36773
45 OTHER RECEIPTS	155763	2691	*41	-	*1207	*2251	29589	9218	15031	12760	13839	13354	55761
46 TOTAL DEDUCTIONS	29419115	318474	*232158	*329753	519667	862620	2441416	1938509	2434745	2390680	2149031	1475134	14326926
47 COST OF SALES AND OPERATIONS	23599890	249225	*143195	*229698	414071	695525	1947096	1631321	1995266	2006458	1748566	1183026	11356442
48 COMPENSATION OF OFFICERS	218677	2719	*5266	*21019	*22705	21211	46389	23127	21259	13406	10150	7692	23734
49 REPAIRS	172871	1633	*1509	*4364	*8111	*6384	10015	14690	10917	11294	12700	1505	89749
50 BAD DEBTS	40449	180	*1491	-	*646	*238	13259	2186	7698	4312	3112	1443	5866
51 RENT PAID ON BUSINESS PROPERTY	151286	1248	*17242	*7091	*3921	*5862	8536	10237	7322	10041	15754	10849	53181
52 TAXES PAID	463716	4391	*9806	*9871	11290	14056	37267	23983	31423	27013	28093	18932	247591
53 INTEREST PAID	428798	6088	*7291	*467	*4304	7690	35914	28379	49562	39349	55176	29742	164635
54 CONTRIBUTIONS OR GIFTS	13982	52	*93	*110	*251	*184	863	424	920	1063	547	2481	6992
55 AMORTIZATION	5320	9	-	-	*4	*8	*18	*229	451	222	64	3563	752
56 DEPRECIATION	495569	4674	*5825	*7241	12856	13339	47387	42693	43699	31760	36103	28804	221190
57 DEPLETION	*1036	-	-	-	-	-	*-0	-	28	12	401	253	341
58 ADVERTISING	717495	10729	*170	*2067	*1504	*3372	20342	17872	29940	23239	25570	24868	557820
59 PENSION, PROF SH, STOCK, ANNUITY	146936	1241	*504	*590	*3252	*3897	7820	5716	7907	5909	5619	9049	95432
60 EMPLOYEE BENEFIT PROGRAMS	115711	2453	-	*1814	*1089	*3426	7877	3950	6658	6362	10341	8867	62875
61 NET LOSS, NONCAPITAL ASSETS	5222	9	*108	-	*43	-	*1178	*398	1347	303	157	499	1179
62 OTHER DEDUCTIONS	2842155	33823	*39658	*45421	35621	87428	257456	133302	220348	209935	196676	143562	1438926
63 TOTAL RECEIPTS LESS TOTAL DEDUCTS	921159	4420	*-15666	*9614	20120	-284	26741	81864	63364	66034	28426	86259	550268
64 CONST TAXABLE INC FRM REL FRN CORPS.	40827	-	-	-	-	-	-	-	7	67	3601	1272	35879
65 NET INCOME (LESS DEFICIT), TOTAL	955924	4420	*-15666	*9614	20120	-284	26703	81864	63292	65120	32004	87531	581207
66 NET INCOME, FORMS 1120, F, L, M	1128562	7588	*1776	*12498	*22186	*16208	63707	90006	86101	76242	79615	91429	581207
67 DEFICIT, FORMS 1120, F, L, M	183325	3168	*17442	*2884	*3394	*16492	40185	*10455	26674	11123	47610	3898	-
68 NET INCOME (LESS DEFICIT), F 1120S	*10687	-	-	-	*1328	-	*3181	*2313	3865	-	-	-	-
69 NET INCOME (LESS DEF), 1120-DISC	-	-	-	-	-	-	-	-	-	-	-	-	-
70 STAT SPEC DEDS, F 1120,F,L,M, TOTAL	51707	-0	-	*292	-	-	*450	*10190	657	11408	12660	10015	6035
71 NET OPERATING LOSS DEDUCTION	39906	-	-	-	-	-	*449	*8324	121	10508	11718	8722	65
72 DIVIDENDS RECEIVED DEDUCTION	11303	-0	-	*292	-	-	*1	*1866	537	895	942	1294	5476
73 OTHER	499	-	-	-	-	-	-	-	-	4	-	-	495
74 INCOME SUBJECT TO TAX, TOTAL	1076953	7588	*1776	*12206	*22186	*16208	63257	79816	85445	64866	67001	81431	575172
75 NET L-T CAP GN TAXED AT ALT RATES	52948	26	-	*110	*102	*682	*183	*14433	1801	1765	341	26782	6723
76 INC TAX (BEFORE CRED), TOTAL (TAX I)	472176	3468	*302	*2657	*5581	*4795	25049	33124	38608	29833	30713	33151	264896
77 REG AND ALTERNATIVE TAX (TAX II)	468324	3468	*302	*2365	*5520	*4677	24880	32818	38354	29228	30663	32651	263400
78 TAX FRM RECOMP PRIOR YR INV CR	2745	-0	-	*292	*61	*118	169	64	205	605	50	48	1133
79 TAX FRM RECOMP PRIOR YR WIN CR	363	-	-	-	-	-	-	-	-	-	-	-	363
80 ADDITIONAL TAX FOR TAX PREFS	*744	-	-	-	-	-	-	*242	49	-	-	453	-
81 FOREIGN TAX CREDIT	44953	16	-	-	-	-	-	*2	40	95	1097	233	43470
82 U.S. POSSESSIONS TAX CREDIT	20733	-	-	-	-	-	702	625	8657	-	-	10749	-
83 INVESTMENT CREDIT	52701	185	*193	*462	*1553	*1092	3091	2133	2337	3349	4650	3353	30303
84 NONREFUND ENERGY CREDIT BEFORE LIM.	2074	-	-	-	-	-	-	1	9	-	804	59	1200
85 WORK INCENTIVE (WIN) CREDIT	103	-	-	-	-	-	-	-0	2	42	36	-	22
86 JOBS CREDIT	3853	-	-	*572	*270	*82	*1153	*434	442	532	234	102	31
87 1979 ESTIMATED TAX PAYMENTS	282016	2606	*330	*1442	*1212	*1863	18820	21131	22192	27556	17245	21936	145683
88 REFUND OF 1979 EST TAX PAYMENTS	12263	996	-	-	-	-	*1801	*282	2591	2132	1039	3422	-
89 REFUNDABLE ENERGY CREDIT	-	-	-	-	-	-	-	-	-	-	-	-	-
90 TRAVEL, ENTERTAINMENT & GIFT EXPENSE	83518	1313	-	*216	*810	*1545	8060	6873	6046	7677	5446	3253	42280
DISTRIBUTIONS TO STOCKHOLDERS:													
91 CASH AND PROPERTY EXC OWN STOCK	235851	598	-	*3040	*5644	*49	6865	4972	14076	7565	6629	5313	181099
92 CORPORATION'S OWN STOCK	37683	656	-	-	-	-	-	-	168	-	-	2000	34859
93 INVEST CREDIT: COST OF PROPERTY	691099	3588	*2148	*11929	*33303	*22372	56234	53010	59300	52278	55015	38621	303300
94 INVESTMENT QUALIFIED FOR CREDIT	635795	3445	*2126	*9214	*25882	*21001	49376	44840	52359	48444	50507	37951	290650
95 TENTATIVE CREDIT	66164	345	*213	*921	*2588	*2100	4938	4484	5247	4844	5051	3795	31638
96 CREDIT CARRYOVER	9524	276	-	-	-	*842	*1930	*736	1517	262	3383	524	54
97 ENERGY INVEST CREDIT: COST OF PROP.	21153	-	-	-	-	-	-	9	377	84	8042	642	12000
98 INVESTMENT QUALIFIED FOR CREDIT	21153	-	-	-	-	-	-	9	377	84	8042	642	12000
99 TOTAL TAX PREFERENCE ITEMS	28208	52	-	-	-	-	*140	*5729	1316	1123	652	10701	8494
100 DISC EXPORT GROSS RECEIPTS	-	-	-	-	-	-	-	-	-	-	-	-	-

* Manufacturing: Food and kindred products: Other food and kindred products

Raking ratio estimation is a method which adjusts the estimated totals of the returns for each post-stratum, namely income-asset major industry class, so that agreement is achieved between the 58 known totals of the major industries and the 9 known totals for the income-asset classes (but not for each of the 522 cells separately). Figure 7 outlines the motivation and the effects of the raking estimation. The last item in that figure indicates that the primary motive is not the appearance of consistency but the expectation that sampling error will be reduced by making the sample estimates agree with known outside information.

Research Plan

Essentially the current method of stratification will be compared to three variations of raking ratio estimation. One method, the usual method, will involve all the post-strata. The other two methods will be limited to post-strata whose sample yield is less than 400 returns or in the other case, the post-strata whose sample yield is less than 200 returns. Post-stratification is thought to be a robust procedure by some in the sense that it gives some insurance against bad samples. [4] The limited versions of raking we will study will shed light on this question since the gains from post-stratification should be greater when the sample size is moderately small. The criterion for comparison will be the relative sampling error (coefficient of variation) which will be estimated from half sample replicates.

We expect that the application of post-stratification methods will give the user and the government better statistics for the dollar spent.

Figure 3

GAINS FROM POST-STRATIFICATION BY INDUSTRY

(Pilot Study)

Item	Reduction in Variance Percent
Inventories	36.7
Business Receipts	26.3
Total Receipts	25.4
Base for Investment Credit	18.9
Depreciation	11.4
Taxable Income	7.9
Capital Gains	- .1
Stockholder's Distribution	- 1.9

NOTE: The items from the corporation tax returns are listed in rank order according to the reduction in variance comparing income-asset stratification with income-asset-industry post-stratification.

Figure 4

DETERMINATION OF SAMPLE STRATA

Size of Total Assets	Asset Code
Under $50,000	1
$50,000 under $100,000	2
$100,000 under $250,000	3
$250,000 under $500,000	4
$500,000 under $1,000,000	5
$1,000,000 under $2,500,000	6
$2,500,000 under $5,000,000	7
$5,000,000 under $10,000,000	8
$10,000,000 under $25,000,000 financial	9

Size of Net Income or Deficit	Income Code
Under $25,000	1
$25,000 under $50,000	2
$50,000 under $100,000	3
$100,000 under $250,000	4
$250,000 under $500,000	5
$500,000 under $1,000,000	6
$1,000,000 under $1,500,000	7
$1,500,000 under $2,500,000	8
$2,500,000 under $5,000,000 financial	9

NOTE: Each sample stratum is labelled 1 through 9. Each return receives an income code and an asset code. The return is assigned to a sample stratum based on the higher of the two codes. For example, sample stratum 2 consist of all returns with asset code 2 and either income code 1 or 2, as well as all returns with asset code 1 and income code 2. Sample stratum 9 is limited to returns in the following financial industries: banks including mutual savings banks and bank holding companies, personal and business credit institutions, other insurance companies, and regulated investment companies.

Figure 5

POPULATION AND SAMPLE COUNTS
BY SAMPLE STRATUM, 1979-1980

Sample Stratum	Population	Sample Count
Part 1. -- 1979		
1	1,064,373	3,583
2	402,114	1,833
3	489,879	3,637
4	283,670	4,734
5	177,821	5,496
6	117,979	9,057
7	39,471	3,837
8	19,653	3,976
9	4,347	2,207
Other Returns	49,839	41,708
TOTAL	2,649,146	80,068
Part 2. -- 1980		
1	1,159,761	4,232
2	432,588	2,349
3	522,736	4,276
4	304,495	5,252
5	192,148	6,253
6	129,074	10,451
7	43,198	4,955
8	23,930	5.406
9	4,739	2,326
Other Returns	54,550	39,965
TOTAL	2,867,219	85,465

Figure 6

COMPARISON OF STRATIFICATION
AND POST-STRATIFICATION

Stratification	Post-Stratification
1. Classify all returns into Income-Asset strata	1. Classify sample returns into Income- Asset- Industry post-strata
2. Select a sample from each stratum	2. ______________
3. Estimate totals for each stratum by weighting up to known counts for the stratum	3. Estimate totals for each post-stratum by weighting up to known counts of major industry
4. Add stratum estimates	4. Add post-stratum estimates

Figure 7

POST-STRATIFICATION BY RAKING
RATIO ESTIMATION

- Stratified estimates of industry groups do not agree with known counts.
- Post-stratified estimates of Income-Asset strata do not agree with known counts.
- Raking ratio estimation adjusts estimates to agree with known counts within a tolerance.
- The raked estimate for each post-stratum is divided by the sample count for the post-stratum to produce a weight. This weight is used in producing all other estimates, for instance, money totals.
- Reduction in sampling error is expected for many estimates using raking ratio estimation.

ACKNOWLEDGMENTS

Technical advice and support by H. Lock Oh and Homer Jones are gratefully acknowledged. Gerald Reiser helped develop the figures and prepare copy which was typed by Sandra Heaney. The research evaluation team for the project consists of Jim Bahnke, Mike Leszcz, Jenny Newport and Dan Rosa.

REFERENCES

[1] Deming, W.E. and Stephan, F.F. On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Tables Are Known, Annals Math. Stat., 1940, Vol. 11, 427-444.

[2] Results of a Study to Improve Sampling Efficiency of Statistics of Corporation Income Westat, Inc., Bethesda, Md. January 1974. (Unpublished)

[3] Internal Revenue Service, Statistics of Income 1978-1979 Corporation Income Tax Returns, Publication 16, U.S. Government Printing Office, Washington, DC 1982.

[4] Holt, D. and Smith, T.M.F. Post Stratification. Journal of the Royal Statistical Society, 1979, A, 142, Part 1, 33-46.