ACCOUNTING FOR DIFFERENCES IN AGGREGATE ESTIMATES AND SIZE DISTRIBUTIONS OF FARM PROPRIETORS' INCOME

Edward C. Budd, The Pennsylvania State University and Bureau of Economic Analysis
Robert Yuskavage, Bureau of Economic Analysis

This paper explores some of the reasons for differences in the reporting of net income from farm self-employment (FSE) in the Bureau of the Census's Current Population Survey (CPS) and farm proprietors' (Schedule F) plus partnership income from tax returns, as estimated in the Statistics of Income (SOI) by the Internal Revenue Service (IRS). These are revealed by differences in both aggregate income estimates and in their size distribution. In general, estimates of farm income derived from the SOI are only a third to a half of the CPS, and the SOI distributions show considerably more inequality than the CPS, particularly in the number and proportion of farm units reporting a break-even (zero income) or a loss. Previous work on farm income estimates has focused on the SOI and the U.S. Department of Agriculture's (USDA) estimates of the net income of farm operators, particularly on the estimates of gross receipts and expenses available from those sources [1]. After adjusting for differences in the coverage of the CPS and IRS income concepts we find that the two estimates are much closer than fractions of 1/3 or 1/2 would suggest, a finding implied in other studies [2]. The USDA estimate, on the other hand, substantially exceeds the other two, even after adjustments to align it more closely with the population coverage and income concept of the CPS or IRS.

The paper is divided into five sections. Section 1 presents a reconciliation of IRS and CPS farm income aggregates. The year-to-year variability in the aggregates is examined briefly in Section 2. Section 3 compares the IRS and CPS farm income size distributions for 1972 and suggests reasons for their differences. Limited evidence on the consistency of reporting in the CPS and on tax returns by individual consumer units is presented in Section 4. A brief summary is given in Section 5.

1. COMPARISONS OF IRS AND CPS AGGREGATES

The IRS farm income estimate derived from the SOI consists of the net income reported by farm sole proprietorships on Schedule F of Form 1040 and by farm partnerships on Form 1065, plus payments to farm partners, which, together with the partner's share of the net income also reported by Form 1040 on Schedule E. The CPS estimate, on the other hand, is a combination of amounts reported in personal interviews by household members (about 80 to 90% of the total) and amounts imputed or allocated to nonreporters whose longest employment that year was farm self-employment (the remaining 10 to 20 percent). Only net FSE income is obtained in the interview.

The left-hand panel of Table 1 compares aggregate farm income estimates for the two sources. SOI farm income averages only 42.4 percent of the CPS from 1966 to 1978; only twice is it over half of the CPS in this 13 year period.

TABLE 1 - CPS AND SOI ESTIMATES OF NET SELF-EMPLOYMENT INCOMES, 1966-1978
(Billions of Dollars)

	Farm			Nonfarm		
Year	CPS	SOI	SOI/ CPS (Pct.)	CPS	SOI	SOI/ CPS (Pct.)
1966	7.8	4.8	61.5	35.0	33.6	96.0
1967	9.8	3.9	39.8	38.3	38.9	101.6
1968	7.7	3.7	48.1	43.8	41.1	93.8
1969	8.5	4.2	49.4	43.6	41.9	96.1
1970	7.9	3.3	41.8	45.3	41.5	91.6
1971	8.4	2.7	32.1	49.7	42.6	85.7
1972	10.6	4.8	45.3	54.2	45.4	83.8
1973	15.7	8.5	54.1	55.8	48.3	86.6
1974	12.8	6.1	47.7	59.5	49.4	83.0
1975	11.9	4.4	37.0	61.5	49.1	79.8
1976	12.6	4.6	36.5	68.0	56.8	83.5
1977	10.2	1.2	11.8	78.3	65.0	83.0
1978	14.6	6.8	46.6	88.6	70.2	79.2

Source: CPS: Bureau of the Census
SOI: Statistics of Income, Business Income Tax Returns, various issues.

The same kind of discrepancy, on the other hand, is not apparent in the reporting of nonfarm self-employment (NFSE) income on tax returns and in the CPS. The right-hand panel of Table 1 shows a similar comparison between the SOI and the CPS for NFSE income. In no year is the SOI as low as 79 percent of the CPS, and averages 88 percent of the CPS over the 13 year period.

These differences raise the question as to whether CPS farm income is too high or the SOI figure is too low relative to each other. While the concept of farm income collected in the CPS appears to be quite similar to that reported on tax returns, judging, at least, by the CPS enumerators' instructions, most of the differences in the left-hand part of Table 1 may well be accounted for by differences in the coverage of the two estimates. The CPS defines FSE income only in rather general terms, and the concept is subject to rather broad interpretation by the respondent, whereas farm income in the IRS has been defined and interpreted more precisely in tax statutes and in both IRS and court decisions. Accordingly, we identify and measure farm-related income in the IRS or other sources which is likely to have been reported in the CPS, but not on farm proprietorship or partnership returns.

Table 2 shows a step-by-step reconcilation procedure for the years 1966 through 1978. Line (1) is the net income of farm sole proprietors and partners (FSP&P), plus payments to partners, as shown in Table 1.

CPS enumerators' instructions state that the net cash (fixed) rent of farm landlords should be entered as net rental income, while landlords' net share (variable) rent is to be reported as FSE income. For tax return purposes, landlords who receive share rent report it either on Schedule F or on Form 4835 (and ultimately on Schedule E), depending on whether they actively participate in the operation of the farm. Not until 1971 were nonparticipating landlords who receive a share rent required to file Form 4835, which is similar to Schedule F in receipt and expense detail; before that date, they were expected to report net rental income on Schedule E. Because time series on the number of Schedule F and Form 4835 returns filed, as well as comparisons with other recipient series, suggest that most such landlords were filing Schedule F's prior to 1971, no adjustment is shown on line (2) for years prior to 1971.

Periodically, IRS conducts an intensive audit study, known as the Taxpayers' Compliance Measurement Program (TCMP), with a sample of tax returns. In both the 1973 and 1976 studies, net farm income from Schedule F of Form 1040 showed an increase of approximately 40 percent, with gross receipts increased and expenses reduced. It seems more likely that CPS respondents report net income amounts that are closer to what they would have reported to IRS had their returns been selected for audit rather than the amounts they actually reported on their returns. For example, SOI NFSE income adjusted for audit is very close to the CPS, averaging only four percent more than the CPS over the same time period. Since a separate audit adjustment is not available for farm partnerships, we assume that the adjustment ratio for sole proprietorships applies to partnerships as well. Line (4), consequently, is 40 percent of line (3).

Receipts from sales of livestock held for draft, breeding, dairy or sporting (DBDS) purposes are reported, not on Schedule F, but on Form 4797, and the net gain from such sales is reported either as a capital gain on Schedule D or an ordinary gain on Form 4797. Since all expenses associated with maintaining the livestock, including depreciation, are reported on Schedule F, the amount of such livestock receipts, minus original cost of purchases (less accumulated depreciation), represents an addition to farmers' net income.

Direct estimates of such gains on livestock sales are available from SOI supplemental reports only for 1962 ($0.718 billion) and 1973 ($1.672 billion). Net gains were 70.6 percent of livestock receipts in 1962 and 81.5 percent in 1973; over half of the returns reported no cost basis for computing the gain. Gains for intervening years were estimated by interpolation, based on the movement in the value of USDA livestock sales. For years subsequent to 1973, the 1973 estimate was extrapolated forward by USDA livestock sales based on the ratio of DBDS livestock gains to USDA livestock sales in 1973. Gains on sales of DBDS livestock from Form 4797 are shown on line (6) of Table 2.

A small amount of farm income is received by persons who for various reasons, legal or illegal, do not file individual tax returns. The only evidence available on nonfilers' income is from the CPS-IRS-SSA 1973 Exact Match (EM) file, which indicates that in 1972 the CPS FSE income of nonfilers was 4.5 percent of total CPS FSE income. In the absence of estimates for any other year, we have used the 1972 percentage to estimate nonfiler income for other years. The estimated net FSE income of nonfilers is shown on line (8).

The adjusted IRS estimate of net farm income is given on line (9) of Table 2. FSE income from the CPS, on line (10), is from Census Bureau tabulations. Line (11) shows the percent the adjusted IRS estimate is of the CPS estimate.

It can be seen from Table 2 that our reconciliation procedure accounts for much, if not most, of the difference between the initial, unadjusted SOI and the CPS. In 1973, the one year for which interpolations or extrapolations of the adjustments were not necessary (except for the nonfiler adjustment), the estimates for the adjusted SOI and the CPS virtually coincide. In 9 of the 13 years, 1966-1978, the adjusted SOI averages 91 percent of the CPS, with no year falling below 83 percent. For the other four--1967, 1970, 1971, and 1977--the ratio of the adjusted SOI to the CPS averages only 67.6 percent, with a low of 58 percent in 1977.

2. ANNUAL VARIABILITY OF FARM INCOME

Another aspect of the difference in the reporting of farm income is the greater year-to-year variability in aggregate income reported on tax returns than in the CPS. Annual percentage changes in the CPS and in the unadjusted and adjusted SOI are given in Table 3. In only three years are the percentage changes in the CPS greater than in either of the two SOI series; in two of these years they move in opposite directions.

One possible explanation for this phenomenon is that some of the respondents in the CPS may be reporting their incomes gross of certain fixed expenses, such as depreciation, interest, and taxes, whereas taxpayers have an incentive to deduct all allowable expenses in reporting net income to IRS. Another, more plausible, hypothesis explaining the relative instability of the SOI estimates compared to the CPS is that farmers are reporting in the CPS their "permanent" incomes, interpreted as some average of incomes realized over immediate past years and the current year, and incomes expected in future years. This latter interpretation is consistent with data on the size distribution of farm income in sections 3 and 4.

3. SIZE DISTRIBUTION OF FARM PROPRIETORS' INCOME

A further problem in the comparison of CPS and IRS farm income estimates is the lack of comparability in their distributions by size of income. In

TABLE 2 - RECONCILATION OF SOI AND CPS ESTIMATES OF NET INCOME FROM FARM SELF-EMPLOYMENT, 1966-1978
(Dollars in Billions)

	Item	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
(1)	Net Income of Farm Proprietors and Partners, SOI	4.783	3.929	3.712	4.155	3.293	2.657	4.828	8.485	6.123	4.448	4.550	1.199	NA
(2)	Net Share Rent, Nonparticipating Landlords	NA	NA	NA	NA	NA	0.225	0.594	1.381	1.851	1.413	1.528	1.467	NA
(3)	Farm Income Before Audit [(1) + (2)]	4.783	3.929	3.712	4.155	3.293	2.882	5.422	9.866	7.974	5.861	6.078	2.666	6.797
(4)	Unreported Farm Income, per Audit	1.913	1.572	1.485	1.662	1.317	1.153	2.169	3.946	3.190	2.344	2.431	1.066	2.719
(5)	Farm Income After Audit [(3) + (4)]	6.696	5.501	5.197	5.817	4.610	4.035	7.591	13.812	11.164	8.205	8.509	3.732	9.516
(6)	Livestock Gains Reported on Form 4797	0.898	0.875	0.918	1.031	1.067	1.107	1.295	1.672	1.505	1.567	1.678	1.726	2.149
(7)	Net Farm Income, SOI Sources [(5) + (6)]	7.594	6.376	6.115	6.848	5.677	5.142	8.886	15.848	12.669	9.972	10.187	5.458	11.665
(8)	Net Farm Income, Nonfilers	0.349	0.439	0.349	0.380	0.356	0.376	0.479	0.706	0.574	0.536	0.565	0.458	0.659
(9)	Net Farm Income, Adjusted, SOI [(7) + (8)]	7.943	6.815	6.464	7.228	6.033	5.518	9.365	16.190	13.243	10.308	10.752	5.916	12.324
(10)	CPS FSE Income	7.760	9.756	7.748	8.455	7.908	8.351	10.645	15.680	12.753	11.906	12.564	10.170	14.647
(11)	Adjusted SOI/CPS [(9) ÷ (10)] x 100	102.4	69.9	83.4	85.5	76.3	66.1	88.0	103.3	103.8	86.6	85.6	58.2	84.1

NA = not applicable or not available.
Sources: see text

TABLE 4 - DISTRIBUTION OF FARM INCOME (GAINS AND LOSSES), BY VIGESILES OF CONSUMER UNITS: MEAN INCOME, RELATIVE MEAN INCOME, AND UPPER INCOME BOUND OF VIGESILE, FOR CURRENT POPULATION SURVEY AND STATISTICS OF INCOME, BEFORE AND AFTER AUDIT, 1972

Percentiles of Consumer Units	Current Population Survey			Statistics of Income					
				Before Audit			After Audit		
	Mean Income	Relative Mean Income	Upper Income Bound	Mean Income	Relative Mean Income	Upper Income Bound	Mean Income	Relative Mean Income	Upper Income Bound
1 - 5	$-4,505	-1.36	$-1,530	$-14,250	-8.59	$-4,600	$-11,970	-4.65	$-3,657
6 - 10	880	-0.27	- 330	- 3,682	-2.22	-2,600	- 2,696	-1.05	-1,636
11 - 15	144	-0.04	1	- 1,970	-1.19	-1,436	- 1,146	-0.45	- 639
16 - 20	1	0.00	1	- 1,138	-0.69	- 883	- 411	-0.16	- 200
21 - 25	41	0.01	190	- 673	-0.41	- 500	- 67	-0.03	- 36
26 - 30	194	0.06	275	- 379	-0.23	- 243	4	a	50
31 - 35	341	0.10	401	- 149	-0.09	- 61	105	0.04	176
36 - 40	522	0.16	600	7	a	60	258	0.10	333
41 - 45	809	0.24	1,000	134	0.08	222	438	0.17	541
46 - 50	1,095	0.33	1,246	298	0.17	352	612	0.24	703
51 - 55	1,532	0.46	1,825	459	0.28	579	875	0.34	1,040
56 - 60	2,039	0.62	2,300	758	0.46	1,000	1,291	0.50	1,514
61 - 65	2,692	0.82	3,000	1,201	0.72	1,414	1,847	0.72	2,123
66 - 70	3,340	1.01	3,900	1,681	1.01	1,974	2,436	0.95	2,880
71 - 75	4,393	1.33	5,000	2,340	1.41	2,840	3,316	1.29	3,736
76 - 80	5,446	1.65	6,239	3,340	2.01	3,917	4,273	1.66	4,839
81 - 85	7,160	2.17	8,000	4,796	2.89	6,001	5,957	2.32	7,208
86 - 90	8,882	2.69	10,000	7,130	4.30	8,743	8,391	3.26	10,158
91 - 95	11,436	3.46	14,000	10,968	6.61	13,979	12,583	4.89	16,037
96 - 100	21,660	6.56	99,000	22,323	13.45	1,131,976	25,340	9.85	1,315,129
96 - 99	17,049	5.16	--	17,771	10.70	--	20,241	7.86	--
100	40,103	12.14	--	40,531	24.43	--	45,735	17.78	--
All units	$3,303	1.00	--	$1,659	1.00	--	$2,572	1.00	--

a = less than 0.005
Source: Bureau of Economic Analysis. Based on tabulations from 1972 Benchmark Income Size Distribution File.

TABLE 3 - PERCENT CHANGE FROM PRECEDING YEAR IN CPS AND SOI FARM INCOME CONCEPTS, 1967-1978

Year	CPS	Unadjusted	Adjusted
1967	+25.7	- 17.9	- 14.2
1968	-20.6	- 6.5	- 4.7
1969	+ 9.1	+ 13.5	+ 11.8
1970	- 6.5	- 21.4	- 15.5
1971	+ 5.6	- 18.2	- 7.8
1972	+27.5	+ 77.8	+ 65.8
1973	+47.3	+ 77.1	+ 67.9
1974	-18.7	- 28.1	- 18.0
1975	- 6.6	- 27.9	- 21.5
1976	+ 5.5	+ 4.6	+ 4.4
1977	-19.1	- 73.9	- 43.3
1978	+44.0	+155.0	+108.3

Source: Calculated from lines (1), (9) and (10) of Table 2.

general, IRS distributions of self-employment incomes, farm or nonfarm, show considerably more relative inequality than CPS distributions. Size distributions of farm income based on CPS and IRS data for 1972 are shown in Table 4.

In order to increase the comparability of these and succeeding distributions, each distribution has been tabulated from BEA's benchmark microdata file of the size distribution of total money income for 1972 [3]. This file is based on the Exact Match (EM) File, which was prepared jointly by the Bureau of the Census and the Office of Research and Statistics (ORS) of the Social Security Administration (SSA); it is the result of an exact match of the CPS with SSA's earnings and beneficiary records and the limited tax information contained in the Individual Master File of IRS. In order to increase the amount and detail of tax return information available, e.g., the size of Schedule F income, the EM was statistically matched by ORS with a subsample of the SOI; the resulting file will hereafter be referred to as the EM-SM file. All the distributions are based on the same recipient unit concept: consumer units (the sum of families and unrelated individuals) rather than tax return units. (There are about 200,000 more tax return units with farm income than consumer units in the EM-SM file.) Comparability among the SOI and CPS distributions has been further increased by including in the SOI distributions the CPS incomes of those CPS units who did not file tax returns, or so-called nonfilers.

Tables 4 and 6 show the dollar mean income and the relative mean income, i.e., the dollar mean divided by the mean of the distribution as a whole, for vigesiles of consumer units (intervals five percentiles wide), as well as a break-out of the top vigesile between the top one percent and the remaining four percentiles. Looked at another way, the relative mean is simply the income share of an interval divided by the size of the interval (in percentiles). It is a convenient way of abstracting from apparent differences in two distributions, occasioned by differences in their dollar means. Finally, the upper income bound is simply the income which separates the given interval from the one immediately above it.

As can be seen in Table 4, the SOI before audit distribution is substantially more unequal than the CPS. In the upper tail of the distribution, for example, the dollar mean incomes of those in the top five percent of the SOI distribution exceed those in the CPS, despite the fact that the overall SOI mean is only half that of the CPS. This fact alone suggests that the problem is not one of farmers' reporting gross income in the CPS and net income to IRS. The major difference between the two distributions is clearly in the number and size of loss incomes (Table 5), with 36 percent showing a loss in the SOI compared with only 14 percent in the CPS--21 percent if CPS break-evens are counted as losses rather than gains. (In the CPS, the respondent has the option of reporting "broke even"; such break-even incomes, which are coded as $1, were reported by seven percent of the CPS recipients of farm income. No comparable category exists in the SOI, since net income is the difference between stated receipts and expenses and could only come out to exactly zero or $1 by coincidence. While we suspect that reporting a break-even is a short-cut way of reporting a loss in the CPS without having to report its amount, there is no way of knowing whether the income of the respondent would be positive or negative if he or she were required to make a specific calculation.)

Part of the difference between the two distributions, as well as in the aggregates, can undoubtedly be attributed to the fact that the returns in the SOI sample are unaudited. The effect on the SOI size distribution of correcting each return in the SOI for the results of audit is shown in the right hand panel of Table 4. Space is lacking to describe in detail the methods used to correct the EM-SM file for audit. Based on the relationships shown by the 1973 TCMP between the income reported by the taxpayer and income as corrected by the auditor, gains were increased and losses reduced by selected ratios for most returns. In addition, as can be seen in Table 5, a net of 266,000 consumer units with IRS farm income were changed from a loss to a gain, and for another 69,000 with a loss, the loss was changed to a zero. The final result was a 20 percent increase in gain and a 28 percent reduction in loss income [4].

The effect of the audit adjustment in raising the mean income of and the reducing the degree of inequality in the SOI distribution is evident in Table 4. It is now those in the top 10 percent of the after audit distribution whose incomes exceed those in the identical part of the CPS distribution, rather than those in the top 5 percent. Most of the effect on the relative distribution comes from the reduction in the number and size of losses, with the proportion of those with a loss being reduced from 36 to 27 percent.

One way of determining whether the major difference between the CPS and the SOI distributions is due to the number and size of loss incomes is to exclude loss incomes from the size distributions. The results are presented in Table 6. Because of the uncertainty as to whether break-evens in the CPS should be interpreted as gains or as losses,

TABLE 5 - CONSUMER UNITS WITH FARM INCOME GAIN, BREAK-EVEN, OR LOSS AND AGGREGATE GAIN AND LOSS, CPS AND SOI, BEFORE AND AFTER AUDIT, 1972

Consumer Units (thousands)	Current Population Survey		Statistics of Income			
			Before Audit		After Audit	
	Number	Pct.	Number	Pct.	Number	Pct.
With Gains	2,530	78.7%	2,100	62.2	2,366	71.6
Break-even	231	7.2	54[a]	1.6	54[a]	1.6
With Losses	453	14.1	1,220	36.2	885	26.8
Total Units	3,214	100.0	3,374	100.0	3,305	100.0
Aggregate Income (millions)	**Amount**		**Amount**		**Amount**	
Gains	$11,503		$9,351		$11,194	
Losses	- 889		-3,753		- 2,693	
Total	$10,614		$5,598		$ 8,501	

[a]The number of break-even incomes of CPS nonfilers included in the SOI before and after audit distributions.

Source: Bureau of Economic Analysis. Based on tabulations from 1972 Benchmark Income Size Distribution File.

TABLE 6 - DISTRIBUTION OF FARM INCOME EXCLUDING LOSSES: MEAN INCOMES AND RELATIVE MEAN INCOMES FOR CURRENT POPULATION SURVEY AND STATISTICS OF INCOME, BEFORE AND AFTER AUDIT, 1972

Percentiles of Consumer Units	Current Population Survey				Statistics of Income[1]			
	Gains Only		Gains Plus Breakevens		Before Audit		After Audit	
	Mean Income	Relative Mean	Mean Income	Relative Mean	Mean Income	Relative Mean	Mean Income	Relative Mean
1 - 5	$ 58	0.01	$ 1	a	$ 43	0.01	$ 51	0.01
6 - 10	184	0.04	6	a	130	0.03	138	0.03
11 - 15	299	0.07	106	0.03	232	0.05	253	0.05
16 - 20	433	0.10	248	0.06	318	0.07	372	0.08
21 - 25	582	0.13	368	0.09	431	0.10	518	0.11
26 - 30	842	0.19	531	0.13	568	0.13	631	0.13
31 - 35	1,059	0.23	768	0.18	789	0.18	814	0.17
36 - 40	1,355	0.30	1,033	0.25	1,085	0.24	1,084	0.23
41 - 45	1,806	0.40	1,340	0.32	1,345	0.30	1,426	0.30
46 - 50	2,118	0.47	1,819	0.44	1,650	0.37	1,840	0.39
51 - 55	2,717	0.60	2,185	0.52	2,052	0.46	2,224	0.47
56 - 60	3,195	0.70	2,825	0.68	2,494	0.56	2,798	0.59
61 - 65	3,883	0.85	3,383	0.81	3,155	0.71	3,417	0.72
66 - 70	4,830	1.06	4,278	1.03	3,854	0.87	4,119	0.87
71 - 75	5,676	1.25	5,162	1.24	4,912	1.10	4,973	1.05
76 - 80	7,076	1.56	6,511	1.56	6,340	1.42	6,663	1.41
81 - 85	8,312	1.83	7,878	1.89	8,126	1.83	8,287	1.75
86 - 90	10,079	2.22	9,677	2.32	10,623	2.39	11,087	2.34
91 - 95	12,884	2.83	12,362	2.97	14,096	3.17	15,210	3.22
96 - 100	23,559	5.18	22,841	5.48	26,811	6.02	28,702	6.07
96 - 99	18,371	4.04	17,879	4.29	21,712	4.88	23,180	4.90
100	44,308	9.74	42,689	10.25	47,206	10.60	50,792	10.74
All Units	$4,547	1.00	$4,166	1.00	$4,453	1.00	$4,730	1.00

a = less than 0.005.

[1]Excludes CPS break-even incomes of CPS nonfilers

Source: Bureau of Economic Analysis. Based on tabulations from 1972 Benchmark Income Size Distribution File.

we have included two distributions for the CPS: one for gains only in the first two columns, and one for the sum of gains and break-evens (each break-even being tabulated as $1) in the second two columns; in both SOI distributions, on the other hand, CPS nonfilers of tax returns who reported a break-even in the CPS have been excluded.

The results are of considerable interest. One result is the much closer correspondence among the overall dollar means, when based on positive incomes only. For example, the CPS mean exceeds the before audit SOI mean by only two percent and falls short of the after audit SOI mean by only four percent when the calculations are restricted to those consumer units reporting a gain. Another result is the considerable narrowing of the rather large differences previously noted in the three relative size distributions--before audit SOI, after audit SOI, and the CPS--when the comparisons are restricted to the recipients of positive, or positive and break-even, incomes. The similarity between the two SOI distributions is not surprising, since, as previously noted, the major effect of the audit correction was on the size and proportion of loss incomes. On the other hand, small but important differences remain between the SOI and CPS relative distributions. As one would expect, the SOI distributions still show more inequality than the CPS, with the relative mean incomes in the SOI exceeding those in either of the two CPS distribution for the highest four or five vigesiles, and lying below those in the CPS for the other vigesiles, except the very lowest. In the lower part of the distribution the SOI distributions are closer to the CPS distribution that includes the break-evens than the one that omits them.

The results support an interpretation of the reporting of farm income in the CPS in terms of some form of a permanent or normal income hypothesis. CPS respondents are not as likely to regard a loss as a normal state of affairs and hence tend to report either a small gain or a break-even for the preceding year to the CPS enumerator in March, even though they may be in the process of preparing a tax return to be filed in April that more accurately reflects their actual income (e.g., a loss) for the same year.

4. CONSISTENCY OF REPORTING IN THE CPS AND TO IRS

The foregoing comparisons suggest only that the reporting of positive incomes is more nearly similar in the two sources than is the reporting of losses and possibly break-evens. The distributions compared are, in effect, the row and column totals of a joint distribution or cross-tabulation of CPS and IRS incomes; they tell us little about the degree of consistency in reporting the presence of farm income in either of the two sources, or, if reported in both, the degree of consistency in the amount reported and in its sign.

The extent of consistency in reporting can only be determined from an exact match of CPS respondents with their corresponding tax returns. Unfortunately, the 1972 EM is far from adequate for this purpose. First, the tax return information included in the EM (which is based on the Individual Master File (IMF), not the SOI) is limited to the amount of adjusted gross income (AGI), wages, interest, and dividends in AGI, and to the presence of such tax schedules as E and F, but not the amount of income reported on them. Second, more than half of the "flags" indicating the presence of a Schedule F were lost in the matching process which created the file.

Certain limited tests with the EM file can, however, be made. Out of 698 tax return units with a Schedule F indicator or "flag," 613, or 88 percent of the persons filing them, reported farm income in the CPS, suggesting a rather high degree of consistency between the filing of a Schedule F and the reporting of CPS farm income. Unfortunately, because of the missing farm flags, no conclusions can be drawn about the converse case: the frequency with which those reporting CPS farm income filed a schedule F.

The only possible test of consistency in the reporting of amounts in the two sources in the 1972 EM file is admittedly crude and indirect. It is a matter of arithmetic that the difference between AGI and the sum of wages, interest, and dividends in the IMF must be equal to the algebraic sum of net incomes reported on the various Schedules (C, D, E, and F), other reported income, and adjustments to AGI. If we restrict ourselves to those EM tax filing units who reported the receipt of farm income in the CPS and who did not file Schedules C, D, and E with their tax return and we assume that other income and the various adjustments to AGI are zero or at least small, we can take the difference between AGI and the sum of wages, interest, and dividends as an indicator of, or "proxy" for, the size of schedule F income. The resulting cross-tabulation for these units is shown in Table 7.

TABLE 7 - RELATIONSHIP BETWEEN IMF PROXY AND CPS FSE INCOME FOR TAX FILING UNITS REPORTING RECEIPT OF FSE INCOME IN THE CPS, 1972

Size of IMF Sched. F Proxy	Units with CPS FSE Income (000)	Mean CPS Amount	Mean IMF Proxy Amount	Relative Mean for Gain Incomes: CPS	IMF
Loss	275.4	$ 229	$-1,283	--	--
Zero	39.2	92	0	--	--
750	230.5	816	264	0.21	0.07
750 - 4,249	279.9	2,763	2,178	0.73	0.57
4,250 - 8,249	92.3	5,427	6,012	1.43	1.58
8,250 - 13,249	52.5	8,960	10,456	2.35	2.74
13,250 - 27,249	33.4	14,763	16,919	3.88	4.43
27,250 or more	10.3	22,857	31,956	6.01	8.37
All Units	1,013.5	$2,691	$2,283	--	--
Units with positive IMF Proxy	698.9	3,806	3,816	1.00	1.00

Source: Bureau of Economic Analysis. Tabulated from 1972 CPS-SSA-IRS Exact Match File. See text.

Perhaps the most interesting part of the table is the 275,400 recipient units--over 27 percent of the total--whose tax returns indicate a farm loss. While the average loss reported is in the neighborhood of $1,300, the units filing these returns reported net gains averaging $229 in the CPS! When the loss and zero brackets are excluded from both distributions, the overall means are virtually identical: $3,806 for the CPS and $3,816 for the IMF. When the distributions are limited to those with positive incomes in the IMF, the IMF shows more inequality than the CPS, with the two relative mean income functions intersecting in the neighborhood of the 70th to 75th percentiles, compared with an intersection between the 81st and 85 percentiles for the CPS gains only and the before audit SOI distributions in Table 6. Given the "noise" in the data underlying Table 7, resulting from the absence of farm flags in the EM and the crude nature of the estimate of schedule F income in the IMF, it is indeed surprising that the results of this last test approximate so closely our previous findings on the size distribution of farm proprietors' income.

5. SUMMARY

In this paper we find that much, if not most, of the difference between the aggregate amount of farm proprietors' income reported in the CPS and that reported on tax returns can be accounted for by apparent differences in the coverage of the two estimates; in particular, the omission in schedule F and partnership returns of the net share rent of nonparticipating farm landlords, gains on DBDS livestock and the farm income of persons not filing tax returns, and by the fact that the SOI estimates are not corrected for audit. The greater variability of IRS compared to CPS farm income is more likely the result of farmers' reporting an estimate of their permanent or normal incomes in the CPS, rather than failing to take account of certain fixed expenses in reporting their incomes in the CPS. The former, but not the latter, hypothesis is also consistent with the cross-section data for 1972 on the size distribution of farm proprietors' income.

In general, the IRS farm income distributions show considerably more inequality than the CPS distributions. Despite the lower overall mean income in the IRS as compared with the CPS, the dollar incomes of those in the upper tail of the IRS distributions, whether before or after audit, actually exceed those in the upper tail of the CPS distributions. Differences between the two sets of distributions can be accounted for primarily by the larger proportion and greater size of losses in the various IRS distributions as compared with the CPS. The overall mean incomes and the corresponding size distributions from the two sources resemble each other much more closely when restricted to those recipient units with positive, or positive plus break-even, incomes, although the IRS still shows somewhat more inequality than the CPS when positive incomes alone are considered. The relative mean incomes of those in the top quintile of the IRS distribution appear to lie above those in the top quintile of the CPS, with those in the bottom 75 or 80 percent of the IRS distributions having relative mean incomes lower than those in the corresponding parts of the CPS distributions. These findings are in agreement with the limited tests of the consistency of reporting of farm income based on the 1972 Exact Match File.

ACKNOWLEDGMENTS

Thanks are due to E. F. Denison and R. P. Parker for helpful comments on an earlier draft of this paper, and to Jean Salter for her work on the audit correction. The views expressed are those of the authors and do not necessarily reflect the position of the Bureau of Economic Analysis.

NOTES AND REFERENCES

[1] Stocker, F. D., and Ellickson, J. C., "How Fully Do Farmers Report Their Incomes?" National Tax Journal, June 1959, pp. 116-126; Houthakker, H. S., "The Great Farm Tax Mystery," Challenge Magazine, Jan.-Feb. 1967, pp. 12-13, 38-39; Reinsel, E. J., Farm and Off-Farm Income Reported on Federal Tax Returns, Economic Research Service, U.S.D.A., August 1968, pp. 27-33.

[2] E.g., Peach, R. W., "Recommendations for the Content and Design of Income Survey Questions Relating to Farm Self-Employment Income," U.S. Department of Commerce, Bureau of Economic Analysis, unpublished, November 1979.

[3] For a description of this file and the methodology used in its creation, see Budd, E. C., and Salter, J. K., "Supplementing Household Survey Estimates of Income Distribution with Data from Other Sources: The U.S. Distribution of Total Money Income for 1972," paper presented at the 17th General Conference of the International Association for Research in Income and Wealth, Gouvieux, France, August 1981.

[4] The implied overall correction ratio from the 40 percent used in Section 1, since the audit corrections to the size distributions were made on an individual return basis, not by applying an aggregate correction to all returns. See J. K. Salter, Improving the Quality of Income Data Reported on Field Surveys and on Individual Tax Returns, unpublished Ph.D. dissertation, The Pennsylvania State University, 1980, for a more complete discussion of audit correction methods applied to the 1972 EM-SM file.

DATA BASES
FOR
SMALL BUSINESS ANALYSIS

Selected Contributed Papers

Sponsored by the Section

on

Survey Research Methods

SMALL BUSINESS DATA BASE: PROGRESS AND POTENTIAL

Bruce A. Kirchhoff and David A. Hirschberg
Small Business Administration

SECTION I
INTRODUCTION

The Small Business Administration, Office of Advocacy, is responsible for carrying out a program of research and analysis to facilitate the growth of small business. PL 94-305, which created this office, directed the establishment of a program of economic research and analysis. Our charge involves understanding the impacts of governmental policy on small business growth and describing the condition of small business to governmental agencies and the Congress. Specifically, Sec. 202 of PL 94-305 lists the following functions:

1. Examine the role of small business in the U.S. economy, and the contribution which small business can make in improving competition, increasing economic mobility, restraining inflation, expanding employment, increasing productivity, stimulating innovation, promoting exports
2. Assess effectiveness of Federal assistance programs on small and minority business
3. Measure direct costs and other effects of regulation
4. Determine the impact of the tax structure on small business

More recently, Congress in PL 96-302 reaffirmed the required development of a small business "indicative data base" [Title I, Sec. 100 A(i)(5)], a small business external data base [Title IV, Sec. 401], and added a Presidential annual report to Congress on the status of small business [Title III, Sec. 303].

This latter requirement specifically asks the President to: (1) "present current and historical data," (2) "identify economic trends" and (3) "examine the effects on small business and competition of policies, programs and activities" of various government agencies.

This combination of legislative mandates clearly identifies objectives for the small business data base:

1. Provide a mailing list of small business in the United States.
2. Provide data describing the current condition of small business.
3. Provide descriptive data over time to identify trends.
4. Provide data for policy analysis.

Congress has not specified priorities among these objectives. It has specified in PL 96-302 detailed content of the four objectives itemizing both the economic and demographic data to be enumerated [Title IV] and the government policies it wants analyzed [Title III]. It leaves the source and form of the data base undefined but defines and budgets a separation of data into "indicative" and "external." And, the legislative history clearly specifies that no data collection burden shall be placed upon small business.

Federal Statistical Data

Congress has given us a unique assignment: "Build a data base without collecting any primary data." This is legitimate since our constituency is already burdened with redundant data collection paperwork imposed upon it by the Federal decentralized statistical system. Thus, we must obtain our data from other agencies. We describe below how we hope to do this.

Ideal "External" Micro Data Base

If confidentiality and other constraints were not factors, it is clear what we would obtain from the various Federal agencies for our micro data base. First, we would draw a sample from the Census Bureau's Standard Statistical Establishment List (SSEL) to accurately describe the small business population (probably about 250,000 firms). The SSEL is a comprehensive list describing the legal status and firm relationship of all establishments in the United States. Data would then be matched from the various five year economic censuses to obtain historical information on employment, sales, assets, and other components by firm. This data set would be matched with financial information from the IRS. For corporations the source would be the 1120 forms which, in addition to the tax liability information, provide key balance sheet and profit and loss information. Form 1065 provides partnership information; and data from the IRS 1040, Schedule C, F, and E provide information on proprietors (nonfarm and farm) and partners.

Next, to collect information on the characteristics of workers in these firms--sex, race, age, earnings, and work histories--Social Security data files would be merged. Finally, the

Federal Trade Commission's Quarterly Financial Report data would be merged to obtain current data on firms.

Such an exact match of records would provide researchers with a nearly complete data base as mandated by Congress. Also it would include micro data necessary for analysis of business response to various government tax, expenditure, regulatory and credit policies, as well as problems of inflation, recession, and productivity. However, present confidentiality rules of IRS, Census, SSA, and FTC make it impossible to obtain access to micro data.

Ideal "Indicative" Data Base

Access to the Standard Statistical Establishment List and to IRS Schedules C, E and F would solve our mailing list compilation efforts. These two files comprise a complete enumeration of the business population. Private data sources may approximate this but will never equal it. Most notably, the SSEL contains the enterprise/establishment linkages that are not complete or are not at all identified in other private mailing lists.

From Ideal to Real

Our ultimate goal is to use the Federal statistical system to develop the ideal data bases. It is cost efficient, avoids duplication, and eliminates any additional data collection paper work burden on business firms. There are however many barriers to our goal: (1) data aggregation, (2) restricted access, (3) different tabulation standards, (4) non-comparable reporting units in micro files, (5) time lags in data publication, and (6) incomplete data sets.

These issues are discussed briefly below:

Data Aggregation

The output of the Federal Statistical system consists almost entirely of aggregate and tabulated data. Although neat and tidy from a producer's point of view, it has an important shortcoming for policy analysis. It is not possible to determine if changes in distributions are due to behavioral changes or shifts in the mix of reporting units.

Access Problem

The lack of access to micro data has frustrated our development efforts since we first began. Virtually all agencies have statutes that restrict interagency transfers of micro data; i.e., data about any one individual business. Unless we can develop new legislation we will be left with a congressional mandate to obtain data from agencies that have congressional mandates to refuse our requests.

Federal Data Tabulation Standards

Each agency has adopted different employment, sales, and asset size standards, and these may change over time. Beginning in 1982, we have developed a definition of Statistical Business Size Categories for tabulating Federal statistics.

Noncomparability of Business Reporting Units

For the most part, each Federal agency that is charged with collecting statistics performs its work independently of other agencies. Industrial classification and geographic coding are generally not coordinated among the various statistical agencies. Thus collection methods differ in important ways that lead to major problems in combining data after collection.

Time Lag in Data Availability

Another difficulty in using existing Federal statistics is that availability is possible ony when they are published. This often occurs with a considerable time lag after collection.

Incomplete Data Sets

Last of all, most Federal statistical agencies collect data in accordance with their needs or Congressional authorization. "Size of independently owned business" is a relatively new concept that is not easily extracted from these existing data sets. Most notably, if we attempt to define size as number of employees, many sources of data are incomplete; i.e. without employment.

SECTION II
EXTERNAL DATA BASE - SBA INTERIM MICRO DATA FILES

Congress clearly sees the external data base as:

1. Providing data describing the current condition of small business
2. Providing descriptive data over time to identify trends
3. Providing data for policy analysis

Researchers would summarize these objectives as longitudinal (over time description) and cause/effect analysis. The Federal statistical system is widely respected for its ability to accurately describe business in the United States for over 40 years. But it is frequently criticized for lacking the data necessary to examine cause/effect relationships necessary for policy analysis. The President's Reorganization Project for the Federal Statistical System concluded that one of the major problems with the current statistical system was lack of policy relevance.

Congress has mandated that the small business external data base must be useful for policy analysis; therefore it must consist of micro data. Since micro data is easily aggregated by computer to

provide descriptive statistics, a micro data base maintained over time will adequately fulfill all three objectives.

Federal statistical micro data on business firms is simply not accessible to SBA under current confidentiality restrictions. Collecting our own data would be prohibitively expensive, e.g. the Bureau of the Census has a budget of $70 million for the 1982 Census of Business. Also Congress, in the legislative history of P.L. 96-302, clearly instructs Advocacy to avoid placing additional paperwork burden on business. Thus, we are compelled to use data collected by others.

This requirement of building a data base without actually collecting data represents a unique challenge. We have pursued several directions simultaneously and have gradually evolved these into three developmental categories: the SBA interim micro data base; Federal statistical system micro data; and special data development projects. These efforts are actually interrelated with each other and the Indicative Data Base. We will discuss each as a separate subject in this and the next section.

Design of the SBA Interim Micro Data Base

Congress has identified several uses of the micro data base. Most notable of these is preparation of data and analysis for the President's annual "Report on Small Business and Competition". In addition, a micro data base on business firms will undoubtedly be of interest to a wide variety of policy analysts. Once again, planning and system design requirements dictate that we identify probable users of this data.

The most important of these efforts involves the use of three Dun and Bradstreet files. These files are the cornerstone of the external and indicative data base effort. Dun and Bradstreet offer three separate data files as described below.

Dun's Market Identifier File

The Dun's Market Identifier (DMI) file contains information on business organizations that had financial activity in any one year. Each record in the file contains the following information on an establishment:

1. Dun's number - This is a number assigned by D&B that can be used to merge it with prior year's files.
2. Geographic location - City, county, state, SMSA, and zip code.
3. Year business started.
4. Annual sales volume.
5. Number of employees.
6. Standard Industrial Classification (SIC) and up to four minor SICs.
7. Parent and headquarter city and state.
8. Dun's number of parent and ultimate parent.
9. Subsidiary indicator.
10. Status indicator - Single location, headquarter, establishment, or branch.
11. Manufacturing indicator - Indicates whether or not manufacturing takes place at the location.

Dun's Trend Files

The Dun's Trend file consists of a set of variables for 600,000 firms appended to the DMI file. It includes for 1973 and 1978 the following variables:

1. Percent growth in sales.
2. Percent growth in employment.
3. Base year sales volume.
4. Base year employment.
5. Sales in 1978.
6. Employment in 1978.

Dun's Financial Statement Files

There are two Financial Statement files. The first consists of over 900,000 companies and provides data for one year. A longitudinal file is available for 324,000 companies containing data for at least two years. The variables in these files include:

1. Date of financial statement.
2. SIC numbers.
3. Number of employees.
4. Geographic location.
5. Year started.
6. Current and previous financial indicators (key balance sheet and profit data).
7. Cash.
8. Accounts receivable.
9. Inventory.
10. Notes receivable.
11. Current assets.

Dun's File Development

Dun's files present two important problems: First, the firms in the file are neither a census of all firms in the U.S. nor a random sample. Thus it is necessary to validate or "benchmark" the files against appropriate sources to be sure that the information drawn from the files accurately describes small business in total. Second, the files are not assembled by statistically rigorous data collection procedures, but instead by voluntary cooperation of respondents. Many firms provide incomplete data, and errors arise from a variety of sources. This makes the files "dirty"; some individual firm records contain missing or obviously incorrect data on one or more items. These records must be located, "cleaned," or rejected from the file before it can be validated or used for analysis.

Cleaning Dun's Files

Brookings Institution was contracted to perform this work. Their progress is detailed in a report entitled: "U.S. Establishment and Enterprise Microdata" (unpublished but copies available from SBA). Their work on the DMI file not only met the needs of the micro data base but also the indicative data base. In fact, for reasons dictated by computer processing and data consolidation, the indicative and external data bases are mixed and matched.

Brookings has successfully linked the three files to gain maximum information availability. They have developed a mechanism so establishments (places of business) can be identified with their appropriate enterprises (organizational units defined by ownership control). Basically this means that all of a firm's subsidiaries and branches (which are recorded as separate establishments in the DMI) are identified as belonging to the parent firm. This is necessary since it is size of parent that defines size of business. The indicative data base now contains a list of establishments defined by size of firm based on enterprise employment. Financial Statement file linkage to the DMI means that data on the DMI not on the Financial Statement file is now available and vice versa. In short, the linked files are far closer to meeting the objectives of a business micro data file.

There are still problems to be worked out. Some DMI employment figures were missing or inconsistent. Missing values have been imputed but inconsistencies remain. DMI sales are similarly afflicted. These problems must be studied and corrected.

Summary

For the reasons described, we are focusing our data base development efforts on the SBA Interim Micro Data files. Cleaning, validating, and extending these files longitudinally are now our major current activities. We hope to have a representative sample of 250,000 businesses in a clean, validated, partially longitudinal form ready for descriptive and policy analysis within FY 1982.

Still, the SBA-IMD will never be "finished." Every year new firms enter, old firms exit, and others grow or decline. This information must enter the longitudinal file as it accumulates current business activity in recognition that such activity will soon be history.

SECTION III
EXTERNAL DATA BASE: OTHER PROJECTS

In Section I, we described a multitude of problems associated with building a micro data base from Federal statistics. We are pursuing solutions to these and describe our actions below. Next we discuss data needs that are clearly necessary for policy analysis, but are not specifically identified in legislation.

Incomplete Data Sets

The preferred definition of business size is based upon total firm (enterprise) employment. However many data sets do not include this measure. At present three separate efforts are being made to add employment to existing data sets.

Imputing Employment into IRS Statistics of Income: A major limitation of IRS Statistics of Income (SOI) is that employment is not available. We intend to impute enterprise employment from one or another source onto the IRS micro data. If successful, IRS will retabulate their statistics by employment size, thereby increasing the descriptive information available.

FTC Quarterly Survey: Congress asked the FTC to reduce the paperwork burden it was placing on small business. In response, the FTC has reduced its sample size and simplified its form. As part of the form change, we have asked for collection of employment data. Questionnaires on these changes were sent out to small business leaders who showed no objection to the additional item. If employment is added, the QFR will be much more useful for examining sales, assets, and profits of small business. The FTC plans to ask for employment data beginning in October 1981.

Commercial Loans: Congress has asked the Federal Financial Institutions Examination Council to determine the feasibility of publicly describing depository institutions' commercial loan portfolios by size of business that is served. It is our hope that we can persuade these institutions to collect and routinely report information on the size of business (employment) borrower as part of the Federal statistical publications.

Self-Employment: 1960-75 Micro Data Samples

We have a longitudinal file on sole proprietors. This file is drawn from one made available by the Social Security Administration. Each year a one percent Continuous Work History Sample (CWHS), based on the same ending digits of the social security number, is drawn from individuals who file an IRS Form SE. This is a tax form for proprietors and partners who have earnings of more than $400 and have not paid the maximum social security tax from wage and salaried employment.

Included in each annual file is information on the sex, race, age, industry, county, and earnings of all covered proprietors. This longitudinal file is at the Bureau of Economic Analysis, Department of Commerce. Approximately 60,000 records are available each year. Because of recent interpretations by IRS of the 1976 Tax

Reform Act's confidentiality provisions data since 1975 have not been made available to CWHS users, including BEA. We would hope that this new confidentiality problem is resolved as soon as possible so that up-to-date information becomes available.

This data will allow description and trend analysis for policy purposes of a segment of small business that is not well described in any other Federal statistical program.

Longitudinal File of Workers by Size of Firm

Along with the Department of Labor's Bureau of Labor Statistics and Employment and Training Administration, an effort was completed to establish a longitudinal file of workers, also based on Social Security's 1 percent CWHS. This file contains a longitudinal quarterly earnings history for each job held by a sample of over one million workers. The variables included in the file are sex, race, age, industry, county, quarterly earnings. In addition special tabulations have been prepared which estimate the employment size of each business firm. This will show differences in the characteristics of workers in small and large companies for the first time.

Annual Survey of Manufactures

Richard and Nancy Ruggles have a two year grant to create a ten year longitudinal file of a sample of manufacturing firms in the U.S. Using Census of Manufactures and Survey of Manufactures data, they plan to build a file containing firm by firm micro data for each of ten years.

When complete the micro file created by the Ruggles will not be fully accessible by researchers. The file will be stored on a limited-access computer. Researchers will prepare analytical programs to examine the file, test these on a simulated sample from the file, and, when satisfied, submit these analytical programs to Census. Census will run the programs on the real file, review the results to assure no breach in confidentiality has occurred, and then give the results to the researcher. This form of limited access to micro data is the best Census can agree to under current confidentiality restrictions and is far greater than what is currently available.

IRS Proposed Access

To date no Federal agency has released business micro data, but our negotiations with IRS have led to a preliminary approach to see if sophisticated masking and sampling techniques can adequately ensure confidentiality, especially with small firms. For the publicly traded corporations, tax data are publicly available from the Security and Exchange Commission from 10K reports.

Special Data Development Projects

Congress was thorough in defining what it wants included in the indicative and external data base. However, other data are also required for adequate policy analysis. Thus we have initiated several special projects to develop policy-relevant data.

Summary Tabulation of History from the MIT Data Base

David Birch, Director of MIT's Program on Neighborhood and Regional Change, has worked with the DMI files for over six years. We have taken advantage of his expertise in several ways. Our first step was to request tabulations of base line data on the distribution of firms and establishments by size, by major industry, and by state and Federal region.

Gross Product Originating by Size of Business

This project for the first time provides annual industry estimates of Gross National Product for small and large business. The time series starts in 1955 and ends in 1976. Small business is defined as fewer than 500 employees, and medium and large business is defined as 500 or more employees.

Summary

As described in Section I, we are working towards an "ideal data base" built from Federal statistics. There are many barriers to be crossed as itemized in Section I. We have and are initiating projects to explore these barriers and crossing them one by one. In the meantime we find ourselves developing two data bases at once; our own indicative and SBA-IMD, and the ideal. We cannot meet the intent of P.L. 96-302 without pursuing these related but separate efforts. The cost of developing the ideal has thus far been over $500,000 per year, but as we proceed to surmount barriers, our opportunities for success grow.

*This is a summary of a more detailed paper by the authors available from SBA. Because of space limitations the mailing list project is not discussed.

ISSUES IN DEVELOPING A MICRODATA BASE FOR SMALL BUSINESS RESEARCH

Vito Natrella, Consultant

The Small Business Administration is required by Title II of the Small Business Act to conduct research in the field of small business. As discussed in this session, a number of approaches have been or are being developed in order to provide the data bases needed for this analysis. One of these is a microdata set representative of both small and large businesses.

Using simulation microdata models, analysis could be developed to evaluate the impacts of different policy choices affecting small business such as economic regulation, environmental and health rules, and taxation. A microdata set could also be used to study the determinants of new business ventures and the characteristics of firms going out of business. A file to meet these needs must contain a variety of information. This would include financial data, geographical information and information on employee characteristics. In addition, a basic requirement is company classification according to employment size.

A considerable amount of data exists from various sources involving these aspects. However, they need to be integrated and made internally consistent. Since no one series contains all the information needed, it is necessary to augment and merge records. Depending on the circumstances this can be achieved by means of an identical match, by a synthetic or statistical match or by a multiple regression technique. Problems associated with these techniques for inputting missing data include high cost, inconsistent definitions, sparsity of cases, poor quality, induced biases, and issues of confidentiality.

A Proposal for a Small Business Microdata Base

Initial development of a small business microdata base could be in the direction of a system of three separate but related files based essentially on the IRS Statistics of Income samples: corporate, partnership, and sole proprietorship. The SOI samples have the advantage that they are stratified random samples and contain almost all the needed data. The system would encompass a Company Financial File, an Establishment File and an Employer-Employee File.

The Company Financial File would consist of the SOI sample files augmented by employment data. The SOI files contain complete income account and tax computation items for all forms of business and, in addition, balance sheet items for corporations and partnerships. Geographic designation is included but refers to the central office for large corporations filing consolidated tax returns. Since employment is to be used as the main size classifier, this item must be added from another source. This might be done from the Standard Statistical Establishment List (SSEL) of Census or from the 941 file of IRS. Each of these has problems which will be discussed later. The Company Financial File could be made more useful by updating some industries on the basis of the Federal Trade Commission Quarterly Financial Report.

In order to permit geographical analysis an Establishment File should be set up. This can be done by obtaining data for the establishments of the firms included in SOI samples in the Company Financial File, from the SSEL of Census. Data would cover industry, employment, receipts, geographic location and employment size of the owning enterprise for each establishment.

The third file in the system of microdata files would contain information on the employees of the firms and would be designated as an Employer-Employee File. This would be developed from the Continuous Work History Sample of Social Security Administration associated with the Establishment File. The file would be arranged by employer and contain data for all its employees. This would permit analysis of the employee mix, characteristics and wages for small business.

Access to Microdata

There are really two parts to the problem of access to microdata. One is that, in order to match data from a number of data files, identification is essential if we want to achieve the greatest accuracy. The other involves obtaining access to microdata without identification so that researchers in small business would have the opportunity and flexibility to work directly with the data.

1. Internal Revenue Service

According to Section 6103 of the Internal Revenue Code authorizations are limited, very specific as to conditions, and restricted to particular stated purposes. Paragraph (j) provides for disclosure of tax returns and return information for statistical use. Only four agencies are included in this section to receive authority for specific statistical purposes. They are the Bureau of the Census for all returns and the Bureau of Economic Analysis for corporate returns, to the extent necessary in the structuring of censuses and national economic accounts. The Federal Trade Commission has access to corporate tax returns in order to select its survey sample for the Quarterly Financial Report. The Treasury has access in order to prepare required forecasts, projections and analyses.

These data may be disclosed to others

only in such form which cannot identify, directly or indirectly, a particular taxpayer. The IRS chief counsel has in past years refused to make the business tax models, without direct identification, available since it was felt taxpayers could be indirectly identified by a knowledgeable person.

2. Bureau of the Census

Release of microdata in identified form to government agencies is even more restricted for the Census Bureau than it is for the IRS. No agencies currently have access although legislation has been framed to permit statistical agencies (as defined in the law) to receive names and addresses of businesses together with some data for classification purposes from the Standard Statistical Establishment List. The basic purpose of the SSEL is to provide a frame of businesses from which a stratified sample can be drawn. While the SSEL is on an establishment basis, employment is aggregated according to owning company.

3. Federal Trade Commission

Access to microdata from the Quarterly Financial Report of the Federal Trade Commission is severely restricted by law similar to the case of the Census Bureau. There are issues involved limiting the usefulness of FTC data. The main deficiency is limitation to only four major industries: manufacturing, mining, wholesale and retail trade. Also, the sample is quite small--15,000 corporations in all four industries. This makes for weakness in the estimates by size of company and industry. Except for the large corporations there is little overlap with the SOI sample. If it is possible to add employment to the QFR survey it may be more feasible to simply have FTC prepare current period estimates by size and relating to SOI tabulations.[1]

4. Social Security Administration

Access to data from the Continuous Work History Sample (CWHS) of Social Security is now in a state of suspense. In the past, microdata had been made available in unidentified form for years prior to 1976. However, IRS has objected to making that data available on the grounds that tax return information could be disclosed indirectly.

Comparability of Data

An important requirement for setting up a microdata file is that data from different files be merged. In addition to the access issue, the issue of comparability of data presents problems. In some cases, while nothing can be done to put the data on a comparable basis, it is desirable to be cognizant of the differences. Some of these arise because of differences in timing, levels of consolidation, or accounting methods.

The SOI corporation sample file, considered the basic file for a small business data base, has certain characteristics which differ from those of files to be used for merging and enhancements. The SOI file for any particular year includes data for corporations with fiscal years ending in July through June of the following year. Aggregate financial data are roughly centered on the calendar year, although only 40 percent of all corporations have calendar year accounting periods. However, these account for over 70 percent of total net income. Also, for the most part, corporation returns, Form 1120, are filed on a consolidated basis wnere consolidation includes subsidiaries owned 80 percent or more. Stockholder annual reports, generally the basis for the Dun and Bradstreet Financial-Economics file, consolidate subsidiaries which are over 50 percent owned.

The most important enhancement to the SOI sample file consists of adding employment to the records so that companies can be classified according to number of employees. Data on total employment appear in the IRS Form 941 filed for the first quarter of the calendar year. Of considerable significance is the fact that businesses file Form 941 on a variety of bases. For the most part, data are for each establishment owned by a particular company under separately issued EINs. They are not associated with the EIN of the parent company. Since Form 1120 is filed for a much higher level of aggregation it is usually not possible to locate employment data for many of the multi-establishment firms.

There are a number of partial solutions to this problem currently in progress. IRS is abstracting EINs for subsidiaries of corporations in its tax year 1979 sample. Census' SSEL, which obtains data on employment and payrolls from the IRS 941 file, associates establishments and subsidiaries with owners and parents. This permits aggregation of employment to a company or enterprise level, closer to that used for Form 1120. While these approaches may not give exact employment figures for large multiestablishment firms, they will provide usable employee size classifications.

Accounting methods provide another area of difference between data from tax returns and data from stockholder reports. Methods used in preparing tax returns are geared to minimizing profits so as to minimize tax. Tax laws also provide special incentives to reducing taxes in order to achieve certain economic or social purposes. These may be handled differently according to standard accounting practice. Examples include depreciation, depletion, tax-exempt interest, and installment sales. The main difficulty with the noncomparability due to accounting methods would arise between the SOI

sample files and the Dun and Bradstreet data. These data would also have the problem of different consolidation rules.

Conclusion

A number of approaches are possible to meet the problems raised above. Some are already underway, while others require considerable planning. The approach that is the most feasible but also the most costly is for SBA to contract out with agencies which already have part of the microdata base to augment and enhance their system. That agency would maintain the file, perform the necessary processing, and produce tabulations needed. SBA has contracted with IRS to have the SOI sample files enhanced with employment data from the 941 file. The Bureau of the Census could also be a candidate for this sort of arrangement. Census has access to the SOI sample files and can use the SSEL to obtain employment figures for enhancement. The SSEL has an advantage over the raw 941 file in that establishment and subsidiary data and EINs are grouped together according to ownership so that total company employment can be determined.

A recent opinion by the Office of the Chief Counsel for the IRS may make IRS access to the Census data possible. The new opinion in effect says that the confidentiality provisions of Title 13 (Census) take precedence over Title 26 (IRS) as applied to IRS employees temporarily sworn in as Census agents. IRS may, therefore, be able to enhance its SOI sample files using the superior SSEL employment data for companies.

An approach which might be more satisfactory to SBA is to modify the microdata files so that they could be made available to another government agency. The objection on the part of IRS to making microdata business files available is that even with identity and location removed it may be possible to identify a taxpayer indirectly. After the SOI sample file has been enhanced with an employment classifier IRS would first remove all direct identifiers. It could then use a system of grouping of large corporations within a particular industry. These groups could be as small as three. IRS is currently studying this and other approaches.

As mentioned by Kirchhoff and Hirschberg in a paper included in this volume, legislation is being introduced to provide SBA with access to the Census Standard Statistical Establishment List. Although it may be difficult, the attempt should be made to change Section 6103 of the Internal Revenue Code to permit SBA to have access to IRS data similar to Census. SBA's requirements are fully backed by the Small Business Act and should be accepted if proper safeguards are designed and made part of the law.

NOTE: This paper was developed on the basis of reports prepared by the author for the Small Business Administration.

1. After the SOI sample has been enhanced with employment data.

ASSOCIATING ESTABLISHMENTS INTO ENTERPRISES FOR A MICRODATA FILE OF THE U.S. BUSINESS POPULATION

Catherine Armington and Marjorie Odle, The Brookings Institution

I. Overview

The Brookings Small Business Microdata Project began work in January 1980 with the goal of defining and building a microdata base including all domestic American businesses. This will be used for analysis of the impact of public policy on the small business sector of the U.S. economy. A four-year project was foreseen which involved:

1. Defining the appropriate population and its relation to aggregate measures of business.
2. Integrating available microdata into a large representative sample.
3. Using other microdata sources to enrich the sample data.
4. Establishing the techniques to develop longitudinal data in order to identify changes in the business population, employment and structure.

The basic reporting unit for the data base we are developing is the business establishment, i.e., a single business location with one or more employees, usually with a single product or service. Employment data are available predominantly on an establishment basis. On the other hand, accounting conventions and other administrative procedures dictate that most other business data be reported on an enterprise (firm) basis. Alternative definitions of "small business" abound in public policy research and implementation. Though numerous committees have been formed and studies conducted to establish standards for differentiating small and large businesses, there is as yet no single accepted definition. For the purposes of this paper we shall define small businesses as firms with fewer than 100 employees.

Working on the generally held assumption that most small business firms comprise only a single establishment, the two reporting units (establishment and enterprise) frequently have been considered equivalent for small businesses. Given our stated definition, 278,000 of the 323,000 multi-establishment firms in our data base do qualify as small businesses. On the other hand, however, 11.5 percent of the establishments with fewer than 100 employees, representing 32 percent of the employment of these small establishments, are actually part of complex (multi-establishment) firms which have more than 100 employees. Therefore, even when considering only small businesses, it is necessary to carefully distinguish enterprise basis data from establishment basis data. This is especially important when comparing data for small businesses with data for large businesses. This paper explains how we established the correspondance between establishments and firms within the 1977 USEEM, our business data base, and discusses the implications of some of the new information derived from these associated data.

II. Source of Establishment Data

The 1977 U.S. Establishment and Enterprise Microdata (USEEM) file is based primarily on data from the Dun and Bradstreet Duns Market Identifiers File (DMI) from early 1979. An extract of economic and descriptive data for each establishment was taken from the DMI File, and the range and distribution of all these data were checked. Various types of errors and minor omissions were corrected. Three percent of the records lacked employment data which we subsequently estimated for them, based on medians calculated for SIC classes within each state from the 1977 County Business Patterns data. A complete description of all the changes in the data and the new structure of the file is available. 1/ The data base now includes 4.7 million business establishments with complete reporting of employment figures and industry classification for both the establishment and the firm, and also age, organizational status and geographic data for each establishment. Other data, not directly relevant to this discussion, are also contained in USEEM. 2/

The 1977 USEEM represents the population of domestic U.S. business establishments with employees around the end of 1977. The coverage is somewhat broader than that of the Census Bureau's County Business Patterns or that of the Unemployment Insurance program of the Department of Labor. The USEEM includes some farm establishments, numerous large semi-governmental businesses in transportation, education and health, and some large non-profit organizations, all of which are excluded from those two governmental data sources.

The original DMI File data included indicators of each establishment's organizational status (e.g., single, headquarters, subsidiary, branch) and pointers to higher level establishments in each firm. Our Multi-establishment Enterprise File (Tree File) began as an extract from the DMI File containing records for all the establishments with ownership ties to other establishments (i.e., complex establishments). The purpose of the Tree File was to provide the data necessary for in depth study of the organizational linkages presented in the DMI File. Such study was preliminary to investigation of the consistency in reporting of employment data in these complex organizations.

At the core of any study of employment data contained in the DMI File are two different reported employment figures. One of these records employment in the establishment, and the second is a more inclusive total firm employment figure. The second is reported for all establishments except branches. For single-establishment firms and establishments classified as the ultimate owner of a firm, this total employment figure represents enterprise employment - the consolidated employment for all establishments in that enterprise, including all subsidiaries and branches. Total employment reported for a subsidiary company represents the employment of the subsidiary and any branches it may have. Table 1 below shows the number of establishments and their employees according to their organizational status in the original DMI File and in USEEM.

Table 1
Number of Establishments and Employment
by Organizational Status
Before and After Tree Completion Process

Establishments

	Simple	Top	Subsidiary	Branch	Total
Before	3,345,000	390,000	78,100	456,000	4,269,000
After	3,414,000	323,000	77,700	884,000	4,699,000

Establishment Employment
(in thousands)

	Simple	Top	Subsidiary	Branch	Total
Before	28,900	12,400	4,984	19,800	66,100
After	30,100	11,900	4,975	38,500	85,500

NOTE: Figures are rounded to thousands (establishments) and hundred thousands (employees). Additional significant digits are included if necessary to make clear distinctions.

SOURCE: Version IIA of the interim file (USA3, Table 15) and Version I of USEEM (V4DMU, Table 8).

When we summed up the reported employment figures in the original file, the discrepancy between aggregate establishment employment and aggregate enterprise employment totalled 15 million employees (65 million in establishments vs 80 million in enterprises). This indicated either a large systematic error in employment reporting or deficiencies in establishment coverage. We checked the levels of employment reported for a large sample of firms and found no evidence of over-reporting. In order to analyze and subsequently eliminate the discrepancy, we needed to clarify the organizational status of each establishment and to group the establishments into enterprises. Then we could (a) complete the organizational structure of each enterprise, (b) determine any employment reporting discrepancy within that firm, (c) correct the discrepancy appropriately, and, finally, (d) reaggregate the establishment employment data for each enterprise.

III. Enterprise Structure Data in the DMI File

The Tree File originally included data for over 924,000 complex establishments. Nearly half of these establishments were branches of firms with a headquarters at a different location. Branches are usually secondary locations, frequently with a different activity or product, but wholly owned and consolidated with the headquarters for accounting purposes. A headquarters is the primary establishment in a firm which has branches. Though occupying the same location, different divisions of a company might be identified as separate branch establishments if they have the characteristics of separate businesses. Over 78,000 establishments were subsidiaries, which were separate legal entities with their own accounting system, but were majority-owned by another establishment (the parent). The other 390,000 complex establishments were presumed to be parents and headquarters. These represented both the owners of the branches and subsidiaries on the file, and also some headquarters establishments whose branches were not reported on the DMI File. The DMI File did not include foreign subsidiaries of domestic businesses, and foreign employment was not included in the total employment figures for parent companies. Furthermore, the DMI File did not include records for sales branches of manufacturing firms; however, the employees in these branches were included in total firm employment figures.

The headquarters/branch relationship is relatively simple, involving only two levels of organization. In the branch record, employment was reported for the branch location, and a code was provided to indicate that it was a branch, along with a pointer to the headquarters record. The headquarters record was coded as a headquarters (which implied that it had at least one branch establishment), but there was no indication of how many branches it had, nor were there any pointers to its branches. Each headquarters record reported a figure for its employment at that location (establishment employment), as well as total employment - which should be the sum of its own establishment employment and that of all its branches. If the data were grouped by firm, these employment figures could be compared to see if all branches had been reported for each headquarters. (See Section V.)

The parent/subsidiary relationship is more complex for two reasons. First, both subsidiary establishments and parent establishments may also be headquarters and have branches under them. Second, parents may also be subsidiaries of other parents. Occidental Petroleum, having nine levels of parent/subsidiary relationships, is an extreme example of this organizational complexity. For each subsidiary record, the DMI File reported the employment at that establishment and, if it was also a headquarters, the total employment of itself and any branches. One code in each record in the DMI File indicated whether the establishment was a headquarters, and another code showed whether it was a subsidiary. Each subsidiary had a pointer to its parent, but there was no code to indicate whether it was itself a parent.

In order to deal more efficiently with these multilevel, complex affiliations between establishments, D & B has used the concept of the ultimate owner, the top of each enterprise structure of related establishments. Every establishment that is part of a multi-establishment enterprise in the DMI File, including the top, should have a pointer to the ultimate owner. Using these pointers, we sorted the establishment records into the enterprises to which they belonged and examined their enterprise level employment data.

The Tree File originally had 390,000 ultimate owner records, each representing the top of a complex enterprise. Most were the simplest form of complex enterprise, i.e., the ultimate owner was a headquarters with one or more branches under it. However, some enterprises, such as ITT, had as many as 1200 associated establishments. While relatively few firms had multi-level, complex structures, the greatest complexity was in the largest businesses; therefore, their proper treatment was essential to achieving accuracy in the data.

Due to the coding scheme used in the DMI File, parent records could only be identified as such if they were not also headquarters or subsidiaries. However, as a result of a Dun and Bradstreet editing error, these recognizable non-headquarters, parent records had had their firm employment figures replaced with establishment employment data, leaving us no indication of the real size of the firm. In order to check for internal consistency and to summarize the data for each complex enterprise, our analysis of complex firms had to work from the bottom up to the ultimate owner.

IV. Correction of Inconsistencies within Complex Establishments

Before reorganizing the complex establishment records into family groups for each firm, it was necessary to ensure logical consistency among the organizational indicators and pointers within each establishment record. Nearly 100,000 records showed evidence of incomplete or conflicting indicators and pointers. We analyzed the sources of logical inconsistencies in the complex establishment records and devised conservative correction procedures. In different cases, conservative meant either minimum change, minimum loss of data, or minimum difficulty in future processing.

We initially identified three distinct sources of inconsistencies:

1. Investigator errors in specifying indicator codes or pointers to parents and headquarters, or keypunch errors in transcribing these data.

2. Time lags between the updating of establishment data on the DMI file and the updating of ultimate owner pointers derived from the Duns semi-annual company affiliation update procedure.

3. Past errors in computer programming or operation that were either undetected or uncorrected.

The effects of these errors sometimes interacted to obscure the primary problem.

Algorithms which detected and corrected fifteen types of inconsistencies were used to make organizational pointers consistent with the organizational status codes. Each branch record was required to have legitimate headquarters and ultimate owner pointers (legitimate defined as different from self). About 39,000 records were coded as branches and had legitimate headquarters pointers, but either pointed to themselves or to non-existent records as their ultimate owner. The pointers were corrected for most of these branches. The remainder, which lacked sufficient information to permit correction, were converted to single, non-branches. Similarly, each subsidiary was required to have a legitimate parent pointer. Most inconsistencies at this level were between the parent and ultimate owner pointers.

Establishments which were neither branches nor subsidiaries, but which had ultimate owner pointers, were required to point to themselves as ultimate owners.

V. Analysis and Correction of Enterprise Structure

After the corrections had been applied to achieve internal consistency in the codes and pointers, the establishments on the file were grouped into enterprises. Establishment records were ordered by ultimate owner; branches and subsidiaries were grouped together within each enterprise. The file was then subjected to completeness analysis. The first step was to identify and extract "topless" enterprises on the file. A "topless" enterprise was one with no establishment record on the DMI File corresponding to the ultimate owner pointers in the member establishments. A designated top or ultimate owner would not have been grouped with its subsidiaries and branches if it were, in fact, a subsidiary of another establishment. Erroneous or incomplete reporting in the family members could produce a misplacement of this type. A search was made throughout the tree file for the reported owners, which were found for about 11,000 of the apparently topless enterprises. Their ultimate owner field was corrected to indicate the actual ultimate owner instead of the establishment mistakenly reported to be the ultimate owner.

About 12,000 establishment records in topless enterprises still lacked ultimate owners. Of these, 7,400 establishment records were coded as branches or subsidiaries, but were related to no other record found on the file. These establishment records were grouped by major industrial group (two-digit SIC) and an ultimate owner record was imputed for each of the 72 groups. About 4,500 establishments remained in 891 topless multi-member enterprises. An ultimate owner record was created for each of these topless enterprises. The SIC code assigned to the imputed ultimate owner record was that of the major industry group accounting for the most employment in the family. All the enterprise family groups then had a top establishment; reported, corrected, or imputed.

The second step of the completeness analysis comprised the examination of each enterprise structure and the verification of its ownership linkages. The analysis was done from the bottom up - first for each subsidiary, then for the top level of the enterprise. A check was made to ensure that all establishments pointed to by the branches in the family were present and were marked as headquarters. Each subsidiary was checked to verify that its immediate parent was either another subsidiary in the same family or the top of the family. Any headquarters which had no branches pointing to it and had total employment equal to or lower than establishment employment had its headquarter status revoked. The same rules were applied to the ultimate owner of the family, ensuring that if the top were coded as a headquarters, it had branches pointing to it. A record coded as an ultimate owner which had no branch or subsidiary records pointing to it and which had equivalent or lower total employment than establishment employment was changed into a single establishment firm. As with all single establishment firms, its total employment was set equal to its establishment employment. A total of 68,000 complex establishments were reclassified as single.

VI. Employment Adjustment and Imputation of Branches

Theoretically, employment total in a subsidiary headquarters record should represent the aggregated establishment employment of itself and all its branches. The total employment figure reported for the ultimate owner or top of an enterprise includes all employment of all domestic establishments owned by the top — subsidiaries and branches. The total employment of a non-headquarters subsidiary should represent only that establishment's employment. When there was evidence that these principles were violated, we reconciled the inconsistencies either by adjusting the total employment figures or by imputing an additional branch establishment.

Adjustments to employment total were needed under two circumstances. The first type occured when aggregate establishment employment was greater than reported employment total. In this situation the total employment field was reset to the sum of establishment employment. The second type of adjustment occured when reported total employment was larger than the aggregate establishment employment, but the difference was considered too insignificant to justify the imputation of an additional branch establishment. This small difference could be due to rounding of large employment figures or to updating of employment figures for some, but not all, of the establishments in an enterprise. The employment difference was considered insignificant when any of the following were true:

a) employment difference was two or less,

b) employment difference was less than 10 and total employment was greater than 1,000 or

c) employment difference was less than 100 and total employment was greater than 10,000. Insignificant differences were corrected by resetting the total employment to aggregate establishment employment.

A new branch establishment was imputed when it seemed reasonable to assume that the discrepancy in employment figures arose from the failure to report separate establishment data for some members of the enterprise. This occured whenever employment total was significantly larger than aggregate establishment employment. Imputed branch records were given unique identifying numbers, and the state code, the ultimate DUNS number, and the headquarters DUNS number of the top record in the family (or sub-family). Their SIC code and industry division were specified as those of the rest of the firm as a whole, determined by the rules used by the Census Bureau for classifying enterprise data for County Business Patterns. Establishment employment for the new branch was set equal to the employment discrepancy between the enterprise employment and the aggregate establishment employment, so that it reconciled the two.

These principles for reconciliation of employment data by adjustment or branch imputation were applied on two levels. First, employment reporting was reconciled in subsidiary groups — that is, parts of enterprises consisting of a subsidiary headquarters and its branches. Then the reconciliation for the ultimate owner was done using the same principles on a full enterprise basis.

Application of these principles at both levels increased both the total number of establishments and the aggregate employment levels. At the subsidiary headquarters level, employment was adjusted in 13,000 records, and 18,000 branch records were imputed. Processing on the full enterprise basis, another 115,000 records had their employment adjusted, and 202,000 branch records were imputed.

VII. Refinement of Branch Imputation

The procedure for imputing a branch to each firm whose employment data indicated incomplete reporting of member establishments did not address the question of how many establishments were not reported. Indeed, we studied a sample of firms and were not able to deduce any general rule to estimate from each firm's data how many branches were missing. We did know that sales branches of manufacturing firms were not reported, but we could not generalize about how many sales branches a given firm should have. We considered relating size of imputed branches to size of reported branches for that firm, but that size often seemed ridiculously small and would have ballooned the number of branch establishments enormously.

Inspection of establishment reporting for a sample of firms showed that, as might have been expected, Dun and Bradstreet frequently reported relatively small central administrative offices, while not covering the large productive branch establishments. Thus, a firm with 200,000 employees might have reported fifty branches or subsidiaries with an average employment of two hundred. The 190,000 employees unaccounted for might very well have been in 10 branches with 19,000 employees in each location. It is certainly unlikely that it would really comprise 950 unreported branches with 200 employees each (except perhaps in retail trade or services).

Designing a reasonable scheme for breaking up the imputed branches where appropriate was essential to enhancing the data and preserving the statistical quality of the data. General rules for the refinement of the branch imputation were therefore necessary. The level of employment reported for branches differed considerably by industry division and by total firm size. Working from tabulations of average branch size by enterprise employment size class for each industry division, we estimated equations for branch size as a function of firm size. Using this calculated branch size for disaggregating imputed branches has the advantage of avoiding distortion of the reported establishment size distribution.

This calculated branch size for each firm with an imputed branch was used, not as the actual size of each imputed branch, but to determine the number of branches which should be used to represent the employment otherwise unaccounted for. The number of branches imputed for a firm was determined by dividing the firm's imputed employment figure (as represented by the employment of the single imputed branch) by the branch size calculated for that firm and rounding down to the nearest integer. Thus, no additional branch would be imputed unless the imputed employment figure was at least twice the average branch size for that size firm in that industry division. A limit of one hundred imputed branches per firm was imposed to restrict imputation for the roughly 300 large firms with most of their employment unaccounted for. Firms with total employment less than twenty were limited to a single imputed branch. This technique allocated the original 202,000 imputed branches into 428,000 branches of more appropriate size.

VIII. Linking Enterprise Data to Establishment Data

The final step in the development of enterprise data was to compute the two most commonly used enterprise characteristics — firm industry division and firm employment size class. These two descriptive data items were appended to the establishment data for each member of the firm.

For employment data, such as County Business Patterns, the Census Bureau defines the industry division of an enterprise as that industry division which accounts for the largest portion of the enterprise's payroll. Using employment as a proxy for payroll, we computed enterprise industry division for each firm by summing up establishment employment classified by the industry division of each establishment's primary SIC. The firm's industry division is the one comprising the largest portion of employment. The mining industry provides a good example of the impact of differing definitions of firm industry. If firm industry division were defined as the industry division of the top establishment of the firm, mining would have included 25,331 enterprises, which own 39,885 establishments with 1,808,000 employees. When the Census Bureau's employment based definition was used, many of these enterprises were reclassified as manufacturing, especially the large oil companies whose refining and petrochemical businesses dominate their employment. Additionally, many small enterprises primarily engaged in mining whose tops were in other industries were shifted into this industry division by the application of this definition. A small net increase in the number of enterprises and establishments resulted, but it was accompanied by a dramatic decrease in employment. The Census Bureau's definition gives us 25,396 enterprises with 40,043 establishments and 1,035,000 employees. 3/

Another important enterprise variable is the enterprise employment size class, which we call firm size class. This is simply a coded variable representing the total employment of the firm, which, after completion of the Tree file, is the actual sum of establishment employment in all the member establishments. Having this datum in the record of each establishment belonging to a complex family permits us finally to analyze, easily and efficiently, our entire file of establishment data classified by firm size. It is usually this size class which is relevant to policy analysis. Because we have the completed Tree file and its associated establishment file, we can now compute other enterprise characteristics that might be needed for special analysis.

IX. Special Uses of Establishment Data with Associated Firm Data

The association of accurate firm size and firm industry division data with each establishment record on the 1977 USEEM provides a solid basic data set on American business establishments and firms. Ideally, this procedure would be repeated for data from several other years and a longitudinal file developed. We have already begun work on data for the 1979 USEEM.

The basic data set now available is a unique tool for analysis and for interpolation of data from other sources. Because our establishment population is well defined and is placed in the context of the owning enterprise for each establishment, USEEM provides a basis for comparison of otherwise non-comparable statistics and a framework for disaggregation of aggregated data.

Consider the question of determining the share of small business in total U.S. business, in regional business, or in particular industries. Rather detailed data on employment by employment size class are available from several sources which would be useful for looking at this question. However, most of these sources provide only establishment basis data and have various limitations on their population coverage. For each source, the comparable population in the USEEM can be defined, and factors can be calculated to convert establishment distributions to enterprise distributions, at whatever level of detail is desired.

For instance, special tabulations of 1972 and 1977 Unemployment Insurance data on employment by employment size class by industry were produced for the Small Business Administration (SBA). These tabulations show that net establishment employment growth from 1972 to 1977 in all industries except government and agriculture can be accounted for as follows:

Table 2

1972-1977 Net Growth in Establishment Employment (Employment in Thousands)

Employment Size Class	Employment Growth	Percent of Total Growth
1 - 99	3,807	51.7
100 - 999	2,604	35.4
1000 +	945	12.9
	7,356	100.0

Source: Special tabulation of unpublished Unemployment Insurance data prepared for Office of Advocacy of the SBA in 1980.

The reporting unit for UI is usually an establishment, but we are interested in employment growth distributed by firm size, not establishment size. For this paper's definition of small business, firm employment under 100, we can use data from the USEEM to convert this distribution into one of small versus large firm size.

Table 3 shows the distribution of employment in the USEEM for establishments with fewer than 100 employees by establishment employment size class and by firm employment size class. Establishments that have fewer than 100 employees, which belong to enterprises with more than 100 employees we call pseudo-small. The percentage of small establishments that are actually pseudo-small is surprisingly high for the establishments with between 20 and 100 employees. In Table 3 notice that 32 percent of employment in small establishments was actually in large firms. If we assume that this employment distribution had not changed substantially since 1972 and that the growth rate for small establishments was independent of their ownership, then we can apportion UI's 3.8 million small establishment employment growth by firm size. Thus 32 percent of the growth in small business employment, 1.2 million employees, is attributable to large firms and the remaining 2.6 million is attributable to small firms. The distribution of net growth in employment shown by UI data becomes 35 percent in small firms versus 65 percent in large firms. Similar procedures can be developed to transform other establishment based distributions at any level of disaggregation, for any definition of small business.

The proportion of small establishments which are pseudo-small differs considerably in various industries. Taking one hundred employees again as the upper limit for small business size, Table 4 shows this variation for the nine industry divisions. The three industries whose small establishments are most dominated by large firms are mining (including petroleum industries); transportation, communications and public utilities; and finance, insurance and real estate. In these industries about 20 percent of the small establishments, with nearly 50 percent of the employment in all small establishments, are owned by large firms. Large firms account for about 30 percent of the small establishment employment in manufacturing, in wholesale and retail trade and in services. Even in the industries with the lowest proportions of pseudo-smalls (under four percent), i.e., construction and agriculture, forestries and fisheries, a substantial amount of small establishment employment (14 percent and 17 percent respectively) is controlled by large firms. Any attempt to analyze economic behavior of firms using establishment employment data should take into account these differences.

X. Summary

Completing the establishment-enterprise association in the 1977 USEEM has provided a unique resource for economic research on U.S. business. The comprehensive population coverage of the original DMI File made the effort and cost of correcting errors and inconsistencies and the reconciling of the employment data worthwhile. The procedures outlined in this paper, while having little apparent effect on the aggregate data for firms and their employment, significantly improved the quality of establishment data. We corrected the codes and pointers for over 200,000 complex establishments. This enabled us to identify 12,000 establishments with apparently non-existent owning firms and to create imputed tops to represent their ultimate owners. The reported firm employment figure was corrected for 195,000 top and subsidiary establishments of complex firms. Finally, we imputed over 420,000 branches to 200,000 firms to compensate for the 19 million establishment employees not accounted for.

Now that the problems have been defined and the solutions tested, the process of editing DMI files from other years and reconciling their establishment and firm employment will be considerably easier. Processing of data from other years is necessary for the next level of research data development - longitudinal establishment and enterprise data files which can be used to study business births and failures, divestitures and acquisitions, and enterprise employment changes.

Table 3

Establishments with Fewer Than 100 Employees
by Establishment and Firm Employment Size Class

Number Employees	Pseudo-Small Firm > 100 Employees	True Small Firm < 100 Employees	Total Small	Pseudo-Small/ Total Small (Percent)
0-4	71,000	2,402,800	2,473,900	2.9
5-9	67,500	854,000	921,500	7.3
10-19	132,700	477,800	610,500	21.7
20-49	170,600	251,900	422,500	40.4
50-99	82,500	61,200	143,700	57.4
TOTAL	524,300	4,047,800	4,572,100	11.5

Employment in Establishments
by Establishment and Firm Employment Size Class
(Employment in Thousands)

Number Employees	Pseudo-Small Firm > 100 Employees	True Small Firm < 100 Employees	Total Small	Pseudo-Small/ Total Small (Percent)
0-4	201	5,716	5,916	3.4
5-9	450	5,442	5,892	7.6
10-19	1,900	6,174	8,074	23.5
20-49	5,168	7,114	12,282	42.1
50-99	5,532	3,889	9,422	58.7
TOTAL	13,250	28,335	41,585	31.9

SOURCE: Version I of USEEM (V4IMU, Table 35).

NOTE: All establishment counts are rounded to nearest hundred. All employment figures are rounded to nearest thousand.

Table 4

Establishments with Fewer Than 100 Employees
by Firm Employment Size Class and Industry Division

Industry Division	Pseudo-Small Firm > 100 Employees	True Small Firm < 100 Employees	Total Small	Pseudo-Small/ Total Small (Percent)
FIRE *	95,100	289,900	394,100	24.1
MINING	8,200	29,400	37,600	21.8
TCPU *	39,200	144,900	184,200	21.3
WHLSE TRADE	72,000	428,900	500,900	14.4
MNFG	57,100	345,800	402,900	14.2
SERVICES	92,000	849,800	941,800	9.8
RET TRADE	140,600	1,292,000	1,432,600	9.8
AGRIC	3,700	103,800	107,600	3.4
CONSTR	16,200	554,200	570,400	2.8
TOTAL	524,300	4,047,800	4,572,100	11.5

Employment of Establishments with Fewer than 100 Employees
by Firm Size and Industry Division
(Employment in Thousands)

Industry Division	Pseudo-Small Firm > 100 Employees	True Small Firm < 100 Employees	Total Small	Pseudo-Small/ Total Small (Percent)
FIRE *	1,827	2,014	3,841	47.6
MINING	238	247	485	49.1
TCPU *	1,123	1,164	2,287	49.1
WHLSE TRADE	1,222	3,024	4,246	28.8
MNFG	1,969	4,270	6,239	31.6
SERVICES	2,945	5,854	8,799	33.5
RET TRADE	3,327	8,085	11,412	29.2
AGRIC	111	550	661	16.8
CONSTR	489	3,128	3,617	13.5
TOTAL	13,250	28,335	41,585	31.9

* FIRE = Finance, Insurance and Real Estate
* TCPU = Transportation, Communications, Public Utilities

SOURCE: Version I of USEEM (V4DMU, Table 35).

NOTE: All establishment counts are rounded to the nearest hundred. All employment figures are rounded to nearest thousand.

FOOTNOTES

1. See U.S. Establishment and Enterprise Microdata (USEEM) - Version I: File Description and also Constance Mitchell, "Employment Imputation from County Business Patterns: Methodology and Production Statistics," Working Paper No. 3. Both are available from the Small Business Microdata Project, Economic Studies Program, The Brookings Institution, Washington, D.C.

2. Approximately 85 percent of the firms on the file have gross receipts data. Associated data from other Dun and Bradstreet data files have been linked to the DMI File data in USEEM to provide data on five-year growth of sales and employment for about 24 percent of the firms. Data from D & B's Financial Statement File have been linked with 22 percent of the USEEM firms, making up an associated file that contains eleven balance sheet and income statement variables and up to five years of historical data for sales, profits and net worth.

3. Figures are derived from tabulations of the data in Version I of the interim file: TREE.V3, TREE.V5, and DMISUM.

RECENT TRENDS IN THE DISTRIBUTION OF EMPLOYMENT BY BUSINESS SIZE AND INDUSTRY

Bruce D. Phillips, U.S. Small Business Administration

1. INTRODUCTION

The recent renaissance of interest in the small business sector has come about, in part, because of the acceptance of small establishments as the creator of a majority of new jobs in the United States, even during the recent 1974-1976 recession.[1] And while the size variable still is frequently considered "a matter of indifference" in the literature,[2] its importance is becoming increasingly well documented.[3,4]

Further recent studies on the small business sector indicate that its contribution to overall economic growth is declining in terms of its share of the Nation's GNP, (from 43 percent to 39 percent from 1963 to 1976), despite the growing numbers of small businesses.[5] Among the factors which have been put forth as causing this declining share of GNP have included regulatory policies and tax policies which discriminate against small businesses, the difficulties of smaller businesses in raising capital, and the lack of compensation to small businesses for assuming the risks of innovation, and for the training of workers for larger businesses, among others. Other discriminatory factors contributing to the decline in the share of GNP contributed by small business have included the lack of adequate representation of small business in the federal procurement process.

The list of factors above, while reasonably complete, suffers from a lack of quantification because of the inability to access micro data, and the absence of a pricing mechanism for some of the externalities listed above, such as the cost of assuming the risk of innovation without a guaranteed return. Employment data, however, at least on an aggregate basis, is one statistic which provides some insight as to those areas in which the small business sector may be declining.[6]

While small business' share of total employment has remained virtually constant from 1972-77, a redistribution seems to be occurring away from mining, wholesale-retail trade and services toward the transportation sector, manufacturing and construction. However the small business sector is declining in those industries which have had the fastest growth rates and which also have been the traditional mainstay of small businesses: services and the wholesale/retail sector. We observe that in the fastest growing sectors of the economy (e.g., forestry and agricultural services, coal mining, crude petroleum refining, insurance, and most of the service sectors) the growth in the small business share has been negative. It is highly likely that this negative growth will extend into the future because the non-manufacturing economy is growing at a faster rate than manufacturing, construction, and agriculture.

While the share of small business has not increased in 70 percent of the industries which are growing nationally, we find that in 13 industries whose employment declined between 1972-77, the share of small business increased in 9 (or 69 percent) of them (simple correlation coefficient = -.21, which is close to being significant). Thus, we come to the hypothesis that the small business share has recently increased faster in declining industries than it has in growing industries. Whether, in fact, there is a substitution of small for large business in declining industries will clearly depend upon many factors, some of which are discussed in the next section.

2. HYPOTHESIS TESTS

From the above discussions, the more difficult question is determining what causes the share of small business to change by industry, since it is not highly correlated with employment growth; these factors are discussed below.

A. Profitability (=RP)

From the literature, we learn that economists have long believed that investments in firms where the efficient or optimum scale of production is large yield higher rates of return than where the optimum scale is small.[7] The reason for these differences across industries would appear to be the quasi-monopolistic capital cost barriers to entry which increase with size.[8] Therefore, within a given industry, it is not surprising that profit rates are higher in larger firms, making it more difficult for small firms to attract capital in this inequitable setting.

Has this hypothesized inverse relationship between firm size and profitability changed during the recent past? If anything, it appears to be getting stronger. In a recent study for the Office of Economic Research of the Small Business Administration, Joel Popkin studied changes in profit type return of the small business sector between 1972 and 1976.[9] Popkin's work - a first attempt to study dynamics - derives the share of gross product originating in the small business sector from 1963 to 1976.[10] Part of Popkin's work was concerned with changes in profits during the 1972-76 recession which was concentrated in large companies in durable goods manufacturing industries.

In Popkin's work, when the percentage of profits rose in the construction, transportation, communication, and utility, and service sectors between 1972-76, for large business, it fell in the respective small business sectors. Further, when the share of profits remained constant in the finance, insurance, and real estate industry for large business, it also fell in the small business sector. Thus, while we do not precisely understand how the transmission of industry profits works from large businesses to small businesses, large profits in big corporations may well not translate into profit gains in the small business sector.

Another recent study, based upon Federal Trade Commission data during the 1974-76 recession showed that in non-durable manufacturing particularly, profits rose in large companies (assets greater than $5 million) and declined in small companies.[11] Thus, once again, there may be a shift of profits (and sales) away from small firms during a recession. This needs further testing. Finally, in some rate of return calculations for available 2 digit industries from the IRS' Statistics of Income, we observe that the ratio of the return on equity of small to large companies varies substantially by industry. Further investigation will also need to be done to see if this is a direct result of varying amounts of capital per unit of output (across industries by size of firm).

In addition, when historical micro financial data becomes available in the future from our analysis of the Dun and Bradstreet Financial Statement files, further clarification of the profit relative variable will occur.[12]

B. Business Failures (=BF)

While it is obvious that a large percentage of business failures are normally associated with small businesses, the exact relationship between the distribution of such business failures, and the share of small business, by major industry, is less well known. In general, changes in the share of small business and business failure may vary directly. A good example is found in construction: between 1972 and 1977, the small business share rose 2.5 percentage points, while the business failure share rose 4.2 percentage points (6.4 percent in absolute terms).

There are, however, exceptions to the above generalization. The small business share in manufacturing increased almost 2 percentage points between 1972-1977, while the failure rate declined in both absolute and percentage terms. Therefore, while a positive relationship between probability of failure and size may be found, it is probably not so strong as previously thought. The source for the business failure data by industry is The Business Failure Record from Dun and Bradstreet, 1978.[13]

In another vein, we may note the difference between "measured" business failures - from which creditors lose money - and all other business failures - which may involve (non-public) insolvency, but which are often not recorded in existing statistics. Thus, it is often only the large mature companies - John Argenti's type 3 failures - which make it into the statistics.[14] Type 1 failures - those that never really get off the ground before failing - and type 2 - those companies that rise quickly to meteoric heights and fall just as quickly - often never make it into anyone's list of statistics.

C. Relative Wages (=RW)

For the first time, the Census Bureau has recently combined information from the (1976) Company Organization Survey and the (1977) Economic Census to produce an estimate of payroll per employee for companies of varying size.[15] The published data have been tabulated for 3 company sizes: those with less than 100 employees, those with 100 to 999 employees, and those with 1,000 or more employees. Our hypothesis concerning the wage variable is that the share of small business (by 2 digit SIC) and the relative wage rate vary inversely. That is, as more small companies come to dominate an industry, the wage differential between the small and large company widens.

Let us see why this negative relationship might be true. First of all, in a static situation, consider that small firms are usually price takers and that generally, other things equal, their benefit packages are lower (medical care might not be free in a small firm for example). This will account for a wage differential between small and large firms; how this varies by industry may be a function of such factors as the degree of unionization in small vs. large firms, product differentiation, and product mix within the 2 digit industries which comprise each major 1 digit cluster.[16]

In the transportation sector, the entry of small trucking, airline, and local transportation companies (an increasing share of the market) might also lead to a wider payroll differential between small and large firms. Clearly, however, the state of local labor markets, product elasticities and other factors facing each size firm will indirectly affect the validity of our hypothesis. The wage variable used in our model and our other data are available upon request from the author.

The construction of this variable for empirical testing deserves brief mention. In most 2 digit industries, we were able to construct a wage index of payroll per employee in establishments with less than 100 employees (a small business proxy) relative to payroll per employee in establishments with more than 1000 employees (a large business proxy) or:

$$RW_{(i)}\ (i=ind) = \left[\frac{\text{payroll/employee(establishment}\leq 100)}{\text{payroll/employee(establishment}\geq 1000)}\right]_{(i)}$$

As expected, in 56/68 or 82 percent of the industries for which data was available, this ratio was less than 1. It exceeded 1 mostly in selected mining and service industries.[17]

D. Availability of Capital (=RKL)

In most industries, it is hardly surprising that the capital-to-labor ratio for large firms is bigger than that for small enterprises. However, we hypothesize that the larger the share of small business in a given industry, the wider is the capital-to-labor ratio for small units compared to large firms.[18]

Consider for example, an industry like hotels.

Where there are several larger firms which dominate in specific tourist locations (like a Hilton, Sheraton, or similar chain), there may be pressure upon small business to equip their units similarly. For example consider a computerized reservation system or a cable TV or in-room movies as items a smaller motel may have to offer. But where much of the location or industry is dominated by small firms, much of the additional capital expenditures may be unnecessary. This argument could obviously be applied to many kinds of businesses (fast food franchises, various manufacturing operations, banking, etc.).

The data we are using to attempt to verify our hypothesis comes from the Source Book for Corporations from the Internal Revenue Service.[19] We use corporations with assets between $1 and $5 million to represent small business (although this is a bit high) and corporations with assets between $25-50 million to represent large businesses. The capital stock data are really stocks and bonds (e.g. obligations) issued by the corporation; the proxy we use for labor costs consists of the sum of salaries of officers of the corporation, contributions to pensions and profit sharing plans and other employee benefit programs. (Direct wage and salary information is not available from the Corporation Source Books.)

Therefore the capital-to-labor relative may be defined as:

$$RKL_i = \frac{E_i / (SA+P+PS)_i \text{ assets 1-5m}}{E_i / (SA+P+PS)_i \text{ assets 25-50m}}$$

where: RKL = capital to labor relative, industry i;

E_i = stocks, bonds, other equity obligations issued by the companies in industry i by asset size class;

SA_i = salaries of the officers of the corporation in industry i by asset size class;

P_i = pensions (paid to) employees of the corporations in industry i by asset size class;

PS_i = profit sharing monies paid to employees of corporations in industry i by asset size class.

E. Mergers and Acquisitions (=RMA)

The seemingly obvious hypothesis would be that mergers and acquisitions adversely impact the market share of small businesses because they (sometimes) eliminate locally based jobs and transfer resources to the parent companies. Some observers, however, disagree. George Benston, in a recently published study for the American Enterprise Institute, concluded that in periods of inflation, merger makes the purchase of capital assets cheaper, helps spread the burden of regulatory and payroll taxes more evenly, and encourages the founding of new businesses.[20] Therefore mergers, in this view, are beneficial to the small business sector.

Contrasted with this view, however, David Birch reports that recently acquired establishments (acquired from 1969 to 1976) have higher death rates and higher contraction rates after merger than before merger.[21] In addition, he reports "acquisition does little to mitigate the effects of a recession."[22]

Given these conflicting views, our hypothesis remains that small business employment shares and (increased) mergers and acquisitions are inversely related.

Insufficient data are available to test this hypothesis. We however have used "Mergers and Acquisitions: 1972-1974," a report by the Census Bureau of 6 major industries covered in the 1972 Economic Censuses. In Table 1 of that report are listed the number of establishments acquired by companies with 500 or more employees for the years 1972, 1973, and 1974. We have chosen to use 1974 as the latest available year.

Clearly, one must normalize most data to prepare it for econometric analysis. In this case, we have used the number of establishments (for similar industries) from the 1974 County Business Patterns. Therefore, the testable variable of interest is:

$$\frac{\text{Number of establishments in industry i acquired by companies of 500+ employees, 1974}}{\text{Number of establishments in industry i, 1974}}$$

(As an alternative denominator, we have also used the number of establishments with more than 500 employees.) Of course the expected sign on the variable in the econometric tests below is negative: the larger the acquisition activity on the part of large firms the smaller the expected small business share in that industry.

F. Employment Growth (=EG)

In section I above, we observed that the correlation between employment growth by industry and change in the small business share by industry was negative and insignificant. This is particularly surprising, because recent research has shown that 2/3 of the new jobs created between 1969 and 1976 were in small establishments.[23]

In theory, therefore, one would hypothesize a positive relationship between general employment growth and the small business share by industry. The problem, therefore, is the usual one of trying to answer a micro question with aggregate data: a refined theory of employment growth by establishment size awaits observations by individual firm. In the interim, we observe only a proxy relationship. Perhaps the segregation by nationally growing and declining industries discussed in section III will be more helpful.

G. Tax Variable (TRR)

While it is not clear what the hypothesized relationship between relative tax payments[24] and the share of small business employment ought to be, it may be reasonable to assume that when taxes are discriminatory (i.e., small firms pay more than their proportionate shares), the likelihood that a business will fail or have lower profits increases. As documented in a recent study by the Wharton school,[25] small business firms face non-corporate taxes which can be in excess of 50 percent of the cash flow before taxes; for large firms the ratio is about one-third. Therefore the burden of non-corporate taxes is higher, on average, by 1/3 in the small business sector. Included in these taxes are license fees, payroll (FICA) taxes, and unemployment compensation, among others.

The relative tax variable TRR which is used in our modelling efforts is more fully discussed in the econometric sections which follow.

3. ECONOMETRIC ANALYSES

Table 1 and structural equation (1) summarize the above discussion of hypotheses. Essentially, changes in the employment share of small business are expected to vary inversely with each of the variables in Table 1 - except employment change and business failures, which are expected to vary positively with small business shares:

$$(1) \quad SB_i = g\ [EG, RP, BF, RW, RKL, RMA, Z]_i,$$

where: SB_i = small business share in industry i = (employment in establishments of 100 employees or less), 1977; and the other variables are as discussed above and listed in Table 1.

Z_i = a vector of other exogeneous variables (to be discussed below).

Identification and Reduced Forms

In examining structural equation (1), it is possible that simultaneity exists. For example, one might hypothesize that relative profitability (RP) is an endogeneous variable, and should be related to those input factors and demand factors which jointly determine it. For example, profit could be a function of relative wages (RW), the relative capital-to-labor ratio (RKL), and other exogeneous demand variables[26] which we have not yet specified in (1). Thus,

$$(2) \quad RP_i = h\ [RW, RKL, Z]_i.$$

On a purely arbitrary basis, we shall hypothesize Z to consist of three relative variables. Each of them is defined for firms of $250-500 thousand in assets relative to the same variable for firms of $25-50 million in assets. (These proxy small to large business ratios, and reflect measurable phenomena.)[27] The variables are taxes paid as a fraction of gross receipts (TRR), relative inventories (RINV), and relative cost of goods sold (RCG). Each of these is expected to vary inversely with profitability since they are subtractions from cash on hand.[28] Obviously these variables could have been added as identities, (i.e. profit = receipts less taxes less cost of goods sold) but our profit relative variable (RP) is only a "dummy" variable, and so an identity is not correct. Equation (2) thus becomes:

$$(2A) \quad RP_i = h'\ [RW, RKL, TRR, RINV, RCG]_i.$$

We might also argue that business failures should be endogeneous in equation (1) above, and vary inversely with relative profitability and positively with employment growth (e.g. more new business failures):

$$(3) \quad BF_i = k\ [RP, EG]_i.$$

Thus business failures are a function of demand (EG) and derived demand (RP). Structural equation (1) therefore reduces to:

$$(4) \quad SB_i = g'\ [RP, BF, RMAI]_i,$$

and the order condition for identification is satisfied.

Therefore, the small business share, business failures and relative profitability are endogenous variables, yielding 3 equations with 3 endogenous variables. In summary, SB, RP, and BF are endogenous; EG, RW, RKL, RMA, TRR, RCG, RINV are exogenous.

Related Studies

It is certain that equations (2) - (4) above are structurally incomplete. Profitability of a company, for example, varies by gross sales, tax rates, location, extent of unionization, and a host of other industry-specific factors.[29] In addition, the paucity of data constrains us initially to a cross-section approach for something that by nature is essentially a time series. After these initial tests, we will (in future efforts) specify a time series model for those series for which data are available.[30]

Econometric Results

Equations (2) - (4) above were first estimated (with 2 stage least squares) in three different ways. The first was for all industries combined, the second was for industries which grew faster (slower) than the U.S. average between 1972-77, and the last was for industries in which the small business share grew more quickly (more slowly) than the U.S. average, 1972-77. Because of the disappointing results with 2SLS, the equations were re-estimated with OLS.

In Table 2, we observe that the best OLS all industry equation is the first one listed. Thus, the small business share across all industries rises .195 percent when general business failures rise 1 percent, and falls - .210 percent when relative mergers and acquisitions rise 1 percent. This first equation explained 45 percent of the variation in the small business shares, and confirmed the merger/acquisition and business failure hypotheses discussed above. The elasticities, however, were relatively small.

We observe from Table 2 that the merger and acquisition variable is only significant in fast growing industries - those in which employment growth

between 1972-1977 exceeded the national average - and those industries in which the small business share exceeded the national average.[31] In the first case, we mean industries such as coal mining and petroleum refining, air transportation, transportation services, finance industries excluding insurance, and most of the service industries. In the latter case are the industries already listed plus the addition of farming, wholesale and retail trade, and most of construction. Clearly (in equation 3) the policy relevant observation is that significant merger activity is responsible for an amazing 80 percent of the loss in market share in these fast growth industries.

In those industries, a one percentage point increase in mergers yields a .42 percentage point decline in employment shares.

Once again in Table 2, we observe that in the slower growing industries (most of manufacturing, general building contractors, finance excluding credit, department stores, communication, utilities, transportation excluding air) the merger and acquisition variable RMA1 does not appear to be a significant factor. However, the level of general business failures in this case is positively correlated with increases in the small business share. (We had already observed in part I above that small businesses are growing in industries with below average growth rates.) From these observations, it may be reasonable that the small business share rises in declining industries when the overall business failure rate increases because large corporations sell their unprofitable subsidiaries. In addition, persons with entrepreneurial talent who are forced to leave positions in large business corporations during recessions may start small business because their own personal opportunity costs decline to virtually zero when they become unemployed (e.g., or to the rate of unemployment compensation).

Subsidiary Hypotheses and Results

In equations 2A and 3 above, we had attempted to use the BF (business failure) variable and relative profitability (RP) variables endogenously in our 2SLS model. Despite the relatively poor performance of these equations, we decided to regress each of the dependent variables against the exogenous variables as above to see if any significant relationships emerged from these simple reduced form tests.

In Table 3, we display two OLS equations using BF as the dependent variable, and the tax rate relative variable TRR as the independent variable. Of the two equations shown, we note especially that the TRR variable has a different (and significant) sign in each of the equations. While the equations themselves are barely significant, let us try to understand what they might mean.

In Table 3, the first equation tells us to expect a decrease in business failures of 1.6 percent in traditional small business industries (generally non-manufacturing but with exceptions as noted above) when relative taxes paid by these companies rise 1 percentage point. I interpret this to mean one of two things. First, because small firms face higher tax burdens than large firms (by about 40 percent on average, in these industries), an increase in taxes might be capable of being shifted forward. If this is true, firms in these industries face relatively inelastic demand curves, which of course is good. On the other hand, only firms with positive profits pay taxes, while failing firms do not; therefore this equation might simply indicate an "ability" to pay taxes, regardless of product elasticity. It will be for future research to distinguish between the two.

In the second equation of Table 3, we observe a very elastic increase of 3.2 percent in business failures when relative tax payments rise 1 percentage point. This might mean that in those industries in which the small business share is small (part of mining, manufacturing, insurance, hotels, motion pictures, communication, utilities, etc.[32]) the role of taxes in driving firms out of business is much more critical.[33] Perhaps the reason is that small companies in these industries are price takers, have a small share of the markets in which they operate, face very elastic demand schedules for their products and therefore cannot shift taxes forward.

The policy implication is therefore to concentrate on tax neutrality by size class in these industries first, and study what percentage of total costs are accounted for by taxes in these industries, and the relationship (or percentage) of taxes to other input factor costs.

Relative Profitability

Table 4 indicates the most significant equations using the RP variable in single variable regressions. We note that the mean of the profit variable is negative; therefore a positive coefficient indicates a negative relationship. Several generalizations seem possible from the OLS regressions in Table 4.[34] First, from equations (1) and (2) in Table 4, we observe that mergers and acquisitions generally lower profits in small businesses. Thus, in industries where the small business share exceeds the mean (as in retail trade and services, for example), a 1 percentage point increase in mergers will lower profits about 2 1/2 percent. This may be because the loss of market power causes the demand curve which the firm faces to shift and/or to become more elastic in inflationary times. This same profit loss phenomena accompanying mergers appears in industries also where the small business share is less than the U.S. average, as in manufacturing (loss of -1.5 percent with an increase in mergers of 1 percentage point). Thus, as small business loses market power due to mergers, profits may also be expected to decline simultaneously.

In equations (4) and (5) of Table 4, we observe that in the "all industry" and "slow growth" industry cases a one percentage point increase in

taxes paid by small business relative to large lowers the profit rate by 4 1/2 percentage points (across all industries) and by 3 percentage points in industries with a smaller than average growth rate. We observe as well the lack of significance of this variable in the case of rapidly growing industries, however. Our tentative explanation for this phenomenon is that perhaps taxes can be shifted forward in rapidly growing industries with relatively inelastic demand curves, while taxes cannot easily be passed on in more slowly growing industries which face more elastic demand schedules for their products.

The irony of the above statement is that it is in the most rapidly growing industries in which mergers and profit declines seem to occur; in these industries, however, high taxes may be a less important factor in explaining why businesses fail than in the more mature industries like parts of manufacturing, finance and mining where in many cases small business already has a small market share (e.g., mergers and acquisitions cannot reduce it much more).

In general, the tax rate relative variable (TRR) in some other stepwise equations contributed toward our being able to explain about 80 percent of the total variation in the profit relative (RP) variable. In the case where the small business share exceeded the mean, for example, TRR explained 35 percent or about half the total of 68 percent of explained variation. And all of the equations in which this variable was significant had large negative elasticities associated with them.

Our final observation from Table 4 is associated with equation (3). In that equation, a 1 percentage point increase in capital intensity is associated with an increase in 2.64 percent in the relative profits of small business firms. Thus while a larger share of small firms in an industry is negatively associated with higher capital intensity - as discussed above - equation (3) in Table 4 indicates that the rate of return to small firms is extremely sensitive to increases in the capital - to - labor ratio, as we have crudely measured it. Therefore a long run policy goal - most directly targeted to rapid growth industries - is to raise the flow of capital to these firms.[35]

4. SUMMARY

We began this paper with a group of observations which concluded that the share of small business has been declining in rapidly growing industries, with much of the recent growth during 1972-78 confined to those industries whose growth rates have been below the U.S. average. We then developed an econometric model which investigated several factors which might be responsible for the phenomena observed above.

Although the estimates of the basic model were not as significant as we had hoped, through a combination of additional single equation OLS and stepwise equations we were able to show the extent to which significant merger activity has lowered the small business employment share, particularly in rapidly growing industries. Further, although some of the evidence has been conflicting, we have seen that the higher (non-corporate) taxes paid by small business, relative to large business, are regressive and have led to lower profits and higher business failures. Finally, we have also observed the extent to which more capital is needed for small firms to obtain a larger market share in rapidly growing industries; this was approximated through the use of a capital - to - labor ratio.

We concluded above that recessions increase general business failures and have also noticed a small but significant increase in the small business share in many industries when general business failures rise; once again, the relationships between overall employment growth, the small business share across industries, and general business failures could not be adequately modelled in this paper because the business failure data was not size specific.

Obviously this paper is the start of a much larger research effort. Much retesting and reformulation remains to be done as better time series data becomes available. Among the issues suggested for further study in this paper are the merger and acquisition effect on small business profits and market shares, a reformulation of the tax variable to include federal income taxes, and a study of the effects of specific types of taxes on small businesses, and to obtain business failure data by size of firm.

ACKNOWLEDGEMENTS

I wish to thank Bruce Kirchhoff, Robert Lipsey, Larry Ledebur, David Birch, Vernon Renshaw, David Hirschberg, Charles Ou, and William Whiston for comments on an earlier draft, and Teri Drischler of Informatics for converting the U.I. data to consistent 1972 SIC codes. A preliminary version of this paper was given at the Small Business Research Conference at Bentley College, Waltham, Massachusetts, March 5-7, 1981.

NOTES AND REFERENCES

1 See David L. Birch, "The Job Generation Process," Center for Neighborhood and Regional Change (MIT, 1979); David L. Birch and Susan McCracken, "Corporate Evolution: A Micro-Based Analysis." Prepared under grant #14151 from the U.S. Small Business Administration, draft, January 1981. See also Harvey A. Garn and Larry C. Ledebur, "The Renaissance of Concern for Small Business Enterprise in the United States," Urban Institute (Washington, D.C.), February 1980.

2 Robert E. Lucas, Jr. "On the size distribution of business firms." Bell Journal of Economics, Autumn, 1978, pp. 508-524.

3
Nonna A. Noto and Dennis Zimmerman, "Federal Assistance to Troubled Firms-An Analysis of Business Failure Data" (Library of Congress, CRS) December 1980 draft.

4
See, for example, Gellman Research Associates, "The Relationship Between Concentration and Technological Innovation." January 1981 (Philadelphia, Pennsylvania)

5
Joel Popkin, "Strategy for a Micro Data Base, Final Report," submitted to the Office of Advocacy, Small Business Administration November 1980.

6
The limitations of our major employment data source for small business-the Unemployment Insurance (U.I.) system of the Department of Labor need to be mentioned so as not to create misinterpretations. First, the system includes all private wage and salary workers, and excludes all self-employed persons and government workers. Therefore any implications regarding the growth or decline of small businesses are made without reference to the self-employed. Secondly, regarding the currency of the data, 1977 rather than 1978 has been used because of the lateness of reporting of several states. Lastly, the data is plagued by an SIC coding change after 1975. 1969-72 data are tabulated on 1967 SIC codes; 1975-78 data are tabulated using 1972 SIC codes; but data had to be retabulated using the consistent 1972 SIC codes to make absolute comparisons at the 2 digit level.

7
W.J. Baumol, Business Behavior, Value, and Growth. (New York, 1959). Quoted in Marshall Hall and Leonard Weiss, "Firm Size and Profitability," Review of Economics and Statistics, August 1967, pp. 319-31. The latter contains a bibliography of such studies.

8
Hall and Weiss, Ibid, p. 326. Their study uses data for the 400 top companies.

9
Popkin, op. cit.

10
The five components of gross product originating are employee compensation, capital consumption allowances, indirect business taxes, net interest and profits. The profit share, therefore, is the residual left over after the first four components have been subtracted. Popkin's study defined small businesses as those with less than 500 employees.

11
Meir Tamari, "Monitoring the Behavior of Small Manufacturing Firms in the Recession." Prepared for the SBA Office of Economic Research under a purchase order; progress report March 10, 1981. (The results were not statistically significant in durable manufacturing.)

12
At present, the relationships between the IRS concept of profits as a rate of return on equity and the rate of profit as calculated by Popkin is not known.

13
Unfortunately this data is not tabulated by company size.

14
John Argenti, Corporate Collapse (McGraw-Hill, London, 1976), Chapter 8.

15
These data are tabulated for all employees in the company, regardless of hours or weeks worked.

16
These kinds of offsetting effects are noted by Stanley Masters in "An Inter-Industry Analysis of Wages and Plant Size," Review of Economics and Statistics, 51, August 1969, pp. 340-345. Masters notes that higher quality labor might be needed in a small firm, (leading to higher wages), but such factors as higher capital intensity, more regimentation, and more unionization may also bid the wage up in large firms.

17
The variable was derived from Table 2 of the U.S. Bureau of the Census, County Business Patterns, #77-54: Enterprise Statistics. (GPO, 1979).

18
The author is unaware of any studies which have looked at precisely this formulation.

19
Internal Revenue Service,Source Book, Statistics of Income--1975, Corporation Income Tax Returns, publication 1053, (IRS, 1/79).

20
George J. Benston, Conglomerate Mergers: Causes Consequences, and Remedies (AEI, Washington, D.C., July 1980) pp. 3-16.

21
David L. Birch and Susan MacCracken, op. cit., p. 52.

22
Ibid.

23
David L. Birch, "The Job Generation Process." Center for Neighborhood and Regional Change,

Massachusetts Institute of Technology, 1979.

24
In the model developed below, non-corporate taxes are expressed as a fraction of gross receipts in 2 asset size classes: $1-$5 million, and $25-$50 million. The data is from the Corporation Source Book of the Internal Revenue Service for 1975.

25
Hans Stoll and James Walter, Tax Incentives for Small Business (Heller Small Business Institute Policy Papers), (Wharton, Philadelphia, Pennsylvania), 1980.

26
Preliminary correlations between RKL and RW were tested to avoid collinearity problems; all were negative and insignificant.

27
As discussed in more detail below, these variables come from the IRS' Source Book for Corporations. Equation (2) is more useful in completing the specification of our model, rather than including variables per se to test specific hypotheses. However, due to the structure of the model, that becomes the end result.

28
Relative inventories to a large extent reflect demand conditions. Generally, the larger the inventories held by a company, the lower its current demand, and the lower its profits. For further information see E.S. Mills, Price, Output and Inventory Policy (New York, Wiley, 1962).

These terms are defined in Eric L. Kohler, A Dictionary for Accountants, 5th Edition, (Prentice Hall), 1975.

29
See, for example, George C. Eddy. "The Small Business Owner: What It Takes to Succeed." Texas Business Review, July-August 1979.

30
At this time, RW and RMA are only available at 5 year intervals. BF is produced annually with incomplete coverage and SB is available only for years 1969-1972, and 1975-1978.

31
The growth rates of these industries are available from the author.

32
This list is not meant to be exhaustive; the actual data is available from the author.

33
Tax payments are about 12 percent higher on average, in the small firms than in large firms.

34
Because the RP variable has a negative mean (e.g., negative change in profit relative), all of the signs on the regression coefficients in Table 4 appear to be the opposite of what they really are.

35
The results with the RKL variable were not significant in slow growth industries.

TABLE 1

Summary of Hypotheses to be Tested in National
Small Business Model (Small Business Equation Only)

	Expected Sign	Neumonic
1. Employment Growth	+	EG
2. Relative Profitability (Dummy)	-	RP
3. Business Failures	+	BF
4. Relative Wage	-	RW
5. Relative Capital-To-Labor Ratio	-	RKL
6. Merger-Acquisition Relative	-	RMA

TABLE 2
Significant OLS Regression Coefficients With Small Business Share (SB)
as Dependent Variable a/

Dependent/Independent:

Industry Type	Constants	RMA1	BF	R^{-2}	F
All industries(1)	38.0809	-5.1348 (-.21)	4.3371 (.195)	.4523	13.6
All industries(2)	36.1779		4.8625 (.218)	.2830	16.2
Fast growth(3)	84.8959	-132.5927 (-.42)		.8048	28.9
Slow growth(4)	32.1928		4.6744 (.214)	.3229	12.9
Small Business Share Exceeds Mean(5)	83.3293	-121.5988 (-.12)		.3650	8.6
Small Business Share Less than Mean (6)	24.0630		2.8138 (.104)	.1660	4.4

Note: Elasticities at the respective means of the variables in parenthesis; omitted variables not included in the respective equations.

a/ All variables significant at $\alpha \leq .05$.

TABLE 3
Significant OLS Regression Coefficients With Business Failures (BF) as Dependent Variable a/

Dependent/Independent

Industry Type:	Constants:	TRR	R^2	F
Small Business Share Exceeds Mean	9.2898	-4.5622 (-1.642)	.234	4.6
Small Business Share Less Than Mean	-2.2721	3.1631 (3.190)	.195	4.8

Note: Elasticities in parenthesis; ommited variables not included in the respective equations.

a/ All variables and equations significant at $\alpha \leq .05$

TABLE 4
Significant OLS Regression Coefficients With Profit Relative (RP) as Dependent Variable a/

Dependent/Independent

Industry Type	Constants	RMA1	RMA2	RKL	TRR	R^2	F
Small Business Share Exceeds Mean (1)	-.0426	.5631 (-2.341)				.290	6.1
Small Business Share Less Than Mean (2)	-.0051		.00053 (-1.4571)			.892	223.8
Fast growth industries (3)	-.0104			-.0221 (2.64)		.233	5.5
All industries (4)	-.0709				.0697 (-4.55)	.212	8.1
Slow growth industries (5)	-.0885				.0934 (-2.99)	.303	9.1

a/ All variables and equations significant at $\alpha \leq .05$

Elasticities in parenthesis; omitted variables not included in the respective equations.

DISCUSSION

Norman Frumkin, Office of Management and Budget*

The five papers cover a wide range of activities in the small business data base work, and consequently I have been highly selective in commenting on a few broad issues. [1]

It will probably be two to three years before some substantive analysis for policy questions can be made from the data base. This is a long time, but when we remember that the small business program was started a few decades ago when Truman was President, it doesn't seem so bad. This will be the first time there will be basic trend data organized systematically for assessing the overall dimensions of the relationship and sensitivity of small business to the rest of the economy.

The authors use different definitions of the cutoff size for small business - e.g. firms with less than 100 or 500 employees have been designated as "small" business. I have simply adopted whatever definitions were used in commenting on the individual papers. However, a new statistical standard for business size data which I will briefly note at the end takes a neutral position on small business definitions.

Bruce Kirchhoff and David Hirschberg discuss the Small Business Administration's overall plan for the data base. Their focus is on developing a micro data base, i.e. a longitudinal data file that follows the progress of the same group of firms over time. I agree that being able to observe gross changes in company fortunes ranging from growing to larger businesses to going into bankruptcy gives much better insight into workings of the real world than conventional data which summarize the net end result of these dynamic changes into the shares of business accounted for by firms of various size. The conventional end result data are important, but they do not have the analytic power of the longitudinal information.

The paper raises the anomaly appearing in existing data that the small business sector accounts for a declining share of overall business activity and yet generates a high proportion of total employment growth. While it can be conjectured that this may reflect something real such as shifts of small business to more labor intensive industries, I think it is equally likely that it is caused by quirks in the data. For example, data on business failures are suspect, and better information on firms going out of business could also result in the anomaly's being a statistical illusion. Thus, although new or rapidly growing small firms have spurts of hiring new people, they also are less financially capable of withstanding economic hard times, and if better data on business failures were available, part of the initial increase in employment might be shown to be short-lived.

Catherine Armington and Marjorie Odle describe the actual work in developing the micro business data file. This is done using the Dun and Bradstreet business lists because of the confidentiality of name and address lists used in Government data collection programs. It is complex work involved with tracing the parent firm ownership of branches and subsidiaries and in accounting for substantial differences in employment between the establishment file of 65 million and the enterprise file of 80 million. It is the core of the longitudinal data file and is an impressive effort in processing masses of diverse records into a consistent framework. The file so far has been developed for 1977, and it is expected this experience will make the file development for later years much easier.

I have a question on checking the overall coverage of the Dun and Bradstreet lists. It would be useful to have statistical comparisons of these files with the Internal Revenue Service and Bureau of the Census business size information. At a minimum it would give the user some overall guide to gaps and divergencies from more complete lists, and it also might provide the capability for revising the file.

Joel Popkin developed an annual time series on the proportion of the gross national product (GNP) accounted for by business firms with up to 500 employees. This is consistent with the GNP by industry series of the Department of Commerce and is the first time such information is available. The data development required a considerable amount of piecing together diverse statistics from the Internal Revenue Service (IRS) and Census Bureau. It is a thoughtful work of estimating for missing and inconsistent data. The time series runs from 1955 to 1976, but the most reliable information is for the economic census years of 1958, 1963, and 1972; estimates for 1977 will also be done. The 1967 census year was not included because a key data item on receipts per company from the IRS Statistics of Income was not available in that year.

The estimated long-term trend of the small

business proportion of the GNP declined from 51-52 percent in 1958 and 1963 to 49-48 percent in 1972 and 1976. Depending on the reference points used, the decline ranges from two to four percentage points. However, because of the problems with the data, I would not cite this as evidence of a decline. I think there is enough margin of error to infer that the proportion has been stable at one-half of the GNP. This is a contribution - we now know essentially what it is, and that it is not one-third or three-fourths of the GNP. In terms of additional insights, if more detailed industry data are now available beyond the nine broad industry categories that were published, it would be interesting to see what industry mix shifts between small and larger businesses have occurred within each of the broad categories.

Vito Natrella pointed up the confidentiality restrictions limiting the use of Government data for the micro data file. This is the reason the Small Business Administration has relied so heavily on the Dun and Bradstreet files. Various proposals have been made that would amend existing legislation to give selected agencies access to data with strong safeguards against leaks of individual company data. This would be an advance by having more consistent data among agencies as well as reducing reporting burden. However, it also involves a broader public policy issue which is the perception that regardless of the safeguards there is the potential for misuse. This issue will have to be explained satisfactorily to the Administration and Congress in order to get new legislation.

Bruce Phillips analyzed possible causes of shifts in employment and business activity associated with small business. He made imaginative use of existing data, and drew tentative interpretations on the role of mergers and taxes. Because of severe data limitations, such as reliance on cross-sectional data in the absence of time series and problems noted earlier with business failure data, it would be premature to use the analyses for policy formulations.

Despite the problems with the data, this work is suggestive of future types of analyses, including the use of supplementary information. One variable that would be interesting is the availability of bank credit to small business, which is the subject of a new study by the Federal banking agencies. Thus, I think there will be long-run benefits in starting the analytic work now by focusing the data base development on problem areas and giving a head start to policy analyses when an adequate time series of the micro data file is available.

I will end with one other aspect of the data base development. This is the formulation of comparable business size categories by number of employees, sales or assets for Federal agencies to follow in tabulating business size data. A new Government-wide standard for tabulating statistical data on business size was developed as part of the work of the interagency committees on small business statistics. The standard does not designate size categories as "small," "medium" or "large," but rather provides the basis for uniform tabulations by all agencies and allows the data user to decide which designations are appropriate. It also has the property of reducing distortions in the size distribution of firms due to inflation, which arise from the fact that when firm size is measured by the dollar value of sales or assets, an upward shift from one size category to the next occurs simply because inflation raises the values of sales and assets. Use of an approximately logarithmic scale in which the successive size class intervals increase by an approximately constant factor reduces the tendency for inflation of dollar values to alter distribution shapes. Of course, this problem does not exist when firm size is measured by number of employees. The standard was developed by Jerry Coffey of the Office of Management and Budget*, and is expected to be included with other Government standards for statistical data.

NOTES AND REFERENCES

* Norman Frumkin and Jerry Coffey were with the Department of Commerce when these comments were given.

[1] The fifth paper, by Joel Popkin, entitled "Measuring Gross Product Originating in Small Business: Methodology and Annual Estimates, 1955-1976" (No. PB81 211799) is available from the National Technical Information Service, Sales Desk, 5285 Port Royal Road, Springfield, VA 22161