TAXPAYER DATA USED TO STUDY WAGE PATTERNS BY SEX AND OCCUPATION: 1969, 1974 AND 1979

Patricia Crabbe, Peter Sailer, and Beth Kilss, Internal Revenue Service

The Internal Revenue Service annually publishes statistics based on data from individual income tax returns (Forms 1040 or 1040A). Users of SOI data may be aware that the statistics are not usually classified by sex or occupation of taxpayers. In this paper we present data for Tax Year 1979 by gender and preliminary data on occupation.

Since it is not possible, using the tax returns alone, to distinguish between the incomes and taxes which should be attributed to each spouse separately, another IRS file was used. Studies in which the Form W-2 was used to distinguish the salaries of husbands from those of their wives were conducted for Tax Year 1969 and repeated for Tax Year 1974.

The W-2's that were used for these reports were attached to the 1040 and 1040A's. However, for 1979 a different procedure was used for getting W-2 information. Since the IRS has another file which consists of W-2 data (the Information Returns Processing File), this file was used instead. Since we used this file, the editors at the IRS Service Centers did not have to go through the trouble of editing the W-2's that were attached to the tax returns.

In order to attempt to occupationally code the 1979 Statistics of Income file we needed the industry code for each taxpayer. Since the W-2's have the Employer Identification Number (EIN) listed on them, it was decided to match these EIN's to the Social Security Administration's Employer Identification File. This would then give us an industry code for each taxpayer. A computerized occupation coding dictionary was then developed from the industry codes and the taxpayers written-in entries as described in last years paper [1].

Organizationally, the present paper is divided into three parts. The first section is devoted to a comparison of the data by sex of taxpayer to comparable data for Tax Years 1969 and 1974 from the Statistics of Income series. The second section compares the male and female wages by occupation of taxpayers. The paper concludes with a brief description of some of the proposed future uses of our file. The appendix presents a very brief review of the methods used to occupation-code the SOI file for Tax Year 1979.

COMPARISON OF INCOME DIFFERENCES BY SEX OF TAXPAYER, 1969, 1974, and 1979

As mentioned earlier, it is not an easy matter to break out the income of married individuals as reported on tax returns, since they are in fact combined on joint returns. Luckily, the occupation study necessitated a match to a second file of W-2 records in order to determine each taxpayer's employer. This match allowed us to bring in amounts of salaries for both taxpayers.

The enhanced 1979 SOI file marks the third time that IRS used Forms W-2 to produce statistics on men's and women's incomes (the previous years being 1969 and 1974). Using data from all three years, some interesting observations can be made. To begin with, we looked at the proportion of joint returns that had two earners. In 1969, 46 percent of joint returns were two-earner families. By 1979, this had risen to nearly 53 percent.

Table 1 compares data on these three years for joint returns with two wage earners. It shows only a very moderate shift in the percentage of total income which was contributed by wives of two-earner families. In the first two years, 42 percent of the wives contributed less than 25 percent of the joint income; this declined slightly to 41 percent in 1979. On the other hand, the percentage of wives contributing

Table 1.--Joint Returns with Two Wage Earners, 1969-1979
(Numbers in thousands)

Percent of Returns' Total Wages	Returns with Two Wage-Earners					
	1969		1974		1979	
	Number	Percent	Number	Percent	Number	Percent
Total	17,283	100	19,076	100	20,723	100
Under 25 percent.....	562	3	715	4	1,136	6
25 under 50 percent..	1,740	10	1,977	10	2,314	11
50 under 75 percent..	7,791	45	8,393	44	8,712	42
75 percent or more...	7,190	42	7,992	42	8,560	41

50 percent or more of the joint income rose from 13 percent in 1969 to 17 percent in 1979.

When comparing amounts of salaries and wages of husbands and wives on joint returns (see Figure 1), it is interesting to note that there has been only a gradual trend in the relationship. In 1969, the husbands' share accounted for 83 percent of the total; between 1969 and 1979, the total amount of salaries on joint returns increased from $390 million to $861 million; in the same period, the percentage of the total accounted for by men dropped only slightly, from 83 percent to 80 percent. If we look at the average salary of those spouses who worked, we find that the average salary for women more than doubled during the ten year period, from $3,428 in 1969 to $7,435 in 1979. However, the average working wife earned about 38 percent of what the average working husband earned in 1969, and by 1979, this had risen only to 40 percent [2].

For the sake of completeness, let us look at statistics on non-joint returns (see Figure 2). The husbands share of salaries and wages remained nearly constant during the 10-year period (57 percent in 1969; 56 percent in 1979). Here, of course, the difference between salaries paid to men and salaries paid to women is much less--not only because unmarried women earn more on the average than their working married counterparts, but because unmarried men earn so much less than their married counterparts. Whether this means men tend to be more successful if they get married, or that successful men are more likely to get married than unsuccessful men, is of course an interesting problem in causality we will not get into here. In any case, there is again only the slightest trend towards the equalization of men's and women's salaries among non-joint returns.

COMPARISON OF MALE AND FEMALE WAGES BY OCCUPATION OF TAXPAYER

When we last presented this type of comparison of male and female wages (for Tax Year 1974 [3]), there was not much we could use in our file to analyze the reasons for the differentials. This time, with the addition of occupation and industry codes to the SOI file, the possibilities become quite intriguing. We could only scratch the surface here, but we have found a few interesting facts.

Overall, the average salary earned by women is only about 50 percent of the average salary for men, when both joint and non-joint returns

Figure 1.

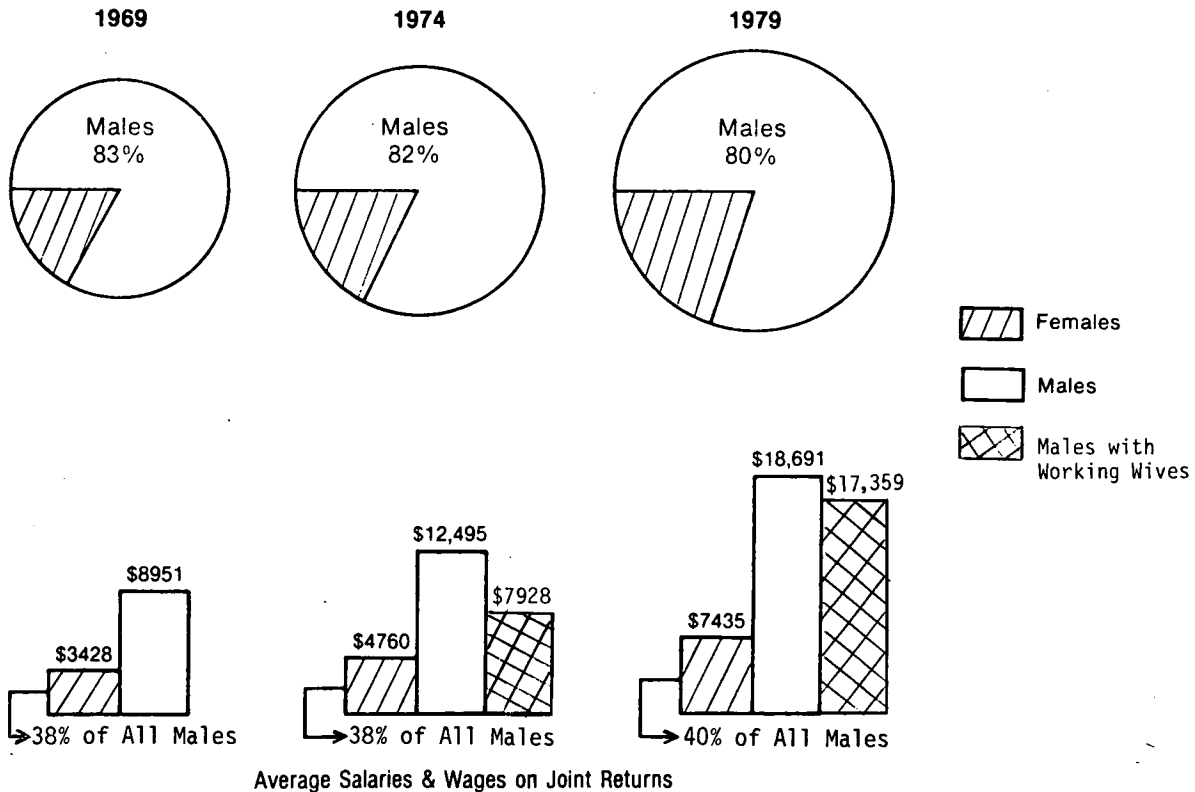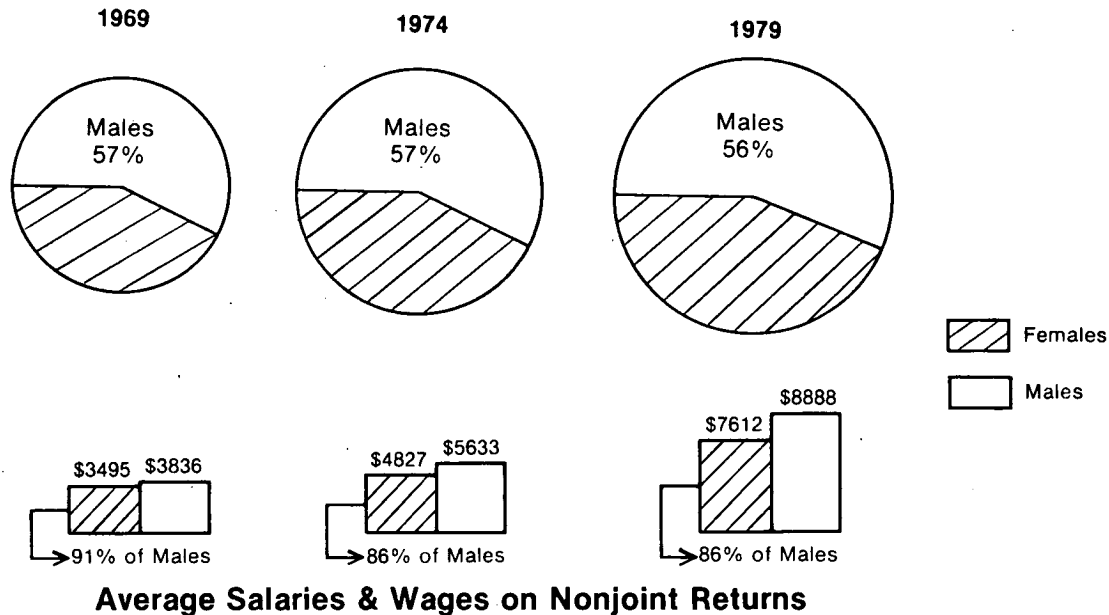## Salaries and Wages on Joint Returns, by Sex

Average Salaries & Wages on Joint Returns

44

Figure 2.

Salaries and Wages on Nonjoint Returns, by Sex

1969   1974   1979

Males 57%

Males 57%

Males 56%

▨ Females

☐ Males

$3495 $3836

$4827 $5633

$7612 $8888

→91% of Males   →86% of Males   →86% of Males

Average Salaries & Wages on Nonjoint Returns

are considered (see Table 2). However, the vast majority of women (about 90 percent) work in occupations where the ratio of female to male wages is higher than 50 percent, although there are virtually no occupations at which women, on the average, earn more than men. A major part of the difference arises from the fact that women are concentrated in some of the lower-paying occupations, such as nursing, where they make up 89 percent of the work force, or elementary and high school teaching, where they make up 66 percent of the work force. On the

other hand, only 26 percent of the administrators in private industry, and only 4 percent of the architects and engineers, were female. Moreover, those women who made it into administration in private industry, on the average, earned only 41 percent of what their male counterparts earned.

This is obviously only a cursory look at the data base we have just created and the analysis we have presented is merely a small indication of some of the interesting questions which might be examined by other researchers. The next and final section of this paper briefly discusses some of the features of the file and also describes several areas we are currently exploring for possible future research.

Table 2.--Salaries from Form W-2 for Selected Occupations

Occupation	Women as Percent of Total Taxpayers	Average Salary	
		Dollars	Women/Men (%)
Total..........	42	11,834	50
OCCUPATIONS IN WHICH WOMEN PREDOMINATE			
Teachers..........	66	12,382	67
Clerical..........	79	9,516	68
Nursing, etc.	89	10,542	80
Retail workers.....	70	6,859	45
OCCUPATIONS IN WHICH MEN PREDOMINATE			
Administrators.....	26	22,576	41
Engineers, Architects, Surveyors..	4	25,912	54
Supervisors in Marketing........	28	14,309	44
Wholesale workers..	15	20,232	46

PLANS FOR THE FUTURE

Since the occupation coding study is still ongoing at IRS, the data we currently have available are preliminary. Our success in developing a computerized occupation dictionary and producing the occupation-coded 1979 SOI file has implications for many other studies currently in the planning stages at IRS. Some features of the SOI data base are worth noting here. Each return record contains either one or two occupation codes (depending on whether it is a joint or non-joint return), one or two industry codes, and all the basic tax return information. Not only salaries and wages, but also pensions and annuities and sole proprietorship income have been distributed between husband and wife on joint returns, based on data available from other administrative records files. Based on the earned income data we had available, we have

also approximated the total income and the total tax burden allocable to each taxpayer. We will be producing a series of Statistics of Income Bulletin articles on the occupations of taxpayers in our 1979 SOI file [4].

While it will not be possible to release all these data elements in the form of a public-use microdata file, due to the disclosure provisions of the Internal Revenue Code, we are working on getting clearance to release a public-use file containing limited income data by sex, marital status, occupation and industry of taxpayer.

In addition, we are already devising some plans to further enhance the file. The major continuing sponsor for this work has been the National Cancer Institute, and their main purpose in sponsoring the work was to obtain a file with which to study occupation-related mortality issues [5]. For this purpose, we are working on plans to run our file up against the National Death Index, to obtain death certificate information, and to code cause-of-death information.

We assume that our data will prove useful, not only to the National Cancer Institute, but also within the Service and the Treasury Department, where they will be available for studying differential tax burdens as well as enforcement related issues. If we are right, we may be able to interest IRS in testing the use of instructions to the taxpayers on how to answer the occupation question, thus improving the quality of the data.

If funds become available to occupation-code another statistical sample, we could code a much larger sample at a much lower cost if the keying of the occupation entry could be integrated into our regular revenue processing system. A sample could be pre-designated by Social Security Number ending digits, and the taxpayer entries keyed every time the new input system (currently being developed for IRS) identifies one of these endings.

We will have an opportunity to test the re-usability of our computerized dictionary very shortly (discussed in detail in [1]). The Department of Defense is mandated by law to conduct an annual study of individuals who have left the Armed Forces, to see how much they are earning on the outside, the better to plan military pay scales. Obviously, they would very much like to see the statistical tabulations we prepare for them classified by occupation, so they know in which areas the differentials are largest [6].

While all this is going on, we are hoping to refine our occupation coding procedures. The Census Bureau has already helped us with a small-scale study of some individual taxpayers whose occupation codes they compared to their own [7]. We are now hoping to arrange a follow-up project where they evaluate our whole dictionary, line-by-line, and compare the entries it would generate to the ones their editors would enter.

We have mentioned just a few of the studies that are possible because of the 1979 Occupation Study. However, the computerized occupation dictionary itself is a valuable product which we hope will be used to enhance the economic data of many other studies here at IRS and elsewhere in the statistical community.

NOTES AND REFERENCES

[1] Crabbe, Patricia, Sailer, Peter, and Kilss, Beth, "Occupation Data from Tax Returns: A Progress Report," 1983 American Statistical Association Proceedings, Section on Survey Research Methods, pp. 312-317.

[2] Much of the increase in the average salary was attributable to inflation. The Consumer Price Index rose from 109.8 in 1969 to 217.4 in 1979 (1967 = 100). Using this index to deflate the 1979 figure, the rise from 1969 to 1979 in terms of 1969 dollars was only from $3,428 to $3,755.

[3] Sailer, Peter. "Using Tax Returns to Study Wage and Taxpaying Patterns by Sex, 1969 and 1974," 1976 American Statistical Association Proceedings, Social Statistics Section, pp. 734-740.

[4] Our current plan, once the data analyses is completed on the occupation distribution, is to publish a series of articles in the Internal Revenue Service's quarterly Statistics of Income Bulletin.

[5] In connection with the Linked Administrative Statistical Sample (LASS) project, the Social Security Administration was one of our earlier financial supporters. The support was to test whether tax returns were codable to the new Standard Occupational Classification System. For more information on the LASS project see Kilss, Beth, Scheuren, Fritz and Buckler, Warren "Goals and Plans for a Linked Administrative Statistical Sample," 1980 American Statistical Association Proceedings, Section on Survey Research Methods, pp. 450-455.

[6] Department of Defense Office of the Secretary of Defense, Fifth Quadrennial Review of Military Compensation, Executive Summary, January 1984.

[7] In an August 1980 Income Survey Development Program (ISDP) Special Frames Test, the Census Bureau included a sample of 800 individuals who were to be included in the 1979 Statistics of Income file. The sample was selected from the 1978 IMF extract which contained mailing addresses at the time of filing (between January 1979 and September 1979). Therefore, there were sample persons not found due to business vs. residence addresses and due to moving between 1979 and August 1980. Results from this study were discussed in last year's paper (See [1]).

APPENDIX-- COMPARISON OF CENSUS AND IRS OCCUPATION-CODING METHODS

The Internal Revenue Service occupations differs in a number of respects from more conventional statistical sources, notably those of the Census Bureau. The most obvious difference is the fact that the tax return contains a little box asking for "Your Occupation" (1 1/2 inches by 1/6 of an inch, as shown in Figure 1), whereas the Census Bureau, in the 1980 decennial Census was able to include two questions on its questionnaire: "What kind of work were you doing?" and "What were your most important activities or duties?" (See Figure 2 on the next page.)

Both systems used the industry code in some instances to generate an occupation code. Here we may have had a slight advantage over Census--at least, our method was less work. Through a series of matches from one administrative records file to another, we were able to access the Social Security Administration's file of all employers in the U.S., which contains a Standard Industrial Classification code for each employer; in the case of multi-unit firms, it even contains, in most cases, separate codes for each unit [1]. At Census, on the other hand, coders had to code the answers each respondent gave to a three-part question on the nature of the firm for which they worked [2].

The Census form contains instructions to the respondent which indicate the level of detail being sought (for example, "nurse" is unacceptable, say "registered nurse."). The total IRS instruction was: "Remember to show your occupations in the spaces in the upper right corner just below the social security number blocks."

Census used a modified form of the Standard Occupational Classification System to code its questionnaires [3]. In the end, we had to modify the SOC slightly for our use as well, but not to the same extent nor always in the same way.

In addition to the Standard Occupational Classification Manual [4], we used coding aids prepared by the Labor Department, notably their Dictionary of Occupational Titles [5]; Census prepared its own coding aids.

The most important difference, however, is that, for the time being, occupation coding is still a manual operation at the Census Bureau. (They are, of course, working on a computerized occupation and industry coding scheme of their own [6].) With the help of our computerized

Figure 1.--1979 Form 1040 Occupation Box

Figure 2.--1980 Census Questionnaire Items 28-30

Facsimile of questionnaire items 28–30	28. A B C
28–30. Current or most recent job activity *Describe clearly this person's chief job activity or business last week. If this person had more than one job, describe the one at which this person worked the most hours. If this person had no job or business last week, give information for last job or business since 1975*	D E F G H J K L M
28. Industry a For whom did this person work? *If now on active duty in the Armed Forces, print "AF" and skip to question 31.* *(Name of company, business, organization, or other employer)*	
b. What kind of business or industry was this? *Describe the activity at location where employed.* *(For example Hospital, newspaper publishing, mail order house, ■ auto engine manufacturing, breakfast cereal manufacturing)*	
c Is this mainly — *(Fill one circle)* Manufacturing ■ Retail trade Wholesale trade Other — *(agriculture, construction, service, government, etc.)*	A F N W
29 Occupation a What kind of work was this person doing? *(For example Registered nurse, personnel manager, supervisor of order department, gasoline engine assembler, grinder operator)*	29 N P Q R S T
b What were this person's most important activities or duties? *(For example Patient care, directing hiring policies, supervising order clerks, assembling engines, operating grinding mill)*	U V W X Y Z

30 Was this person -- *(Fill one circle)*
 Employee of private company, business, or individual for wages, salary, or commissions ■
 	
 Federal government employee
 State government employee
 Local government employee *(city, county, etc.)*
 	
 Self employed in own business, professional practice, or farm —
  Own business not incorporated
  Own business incorporated
 	
 Working without pay in family business or farm	

occupation coding dictionary, we feel that we achieved a high degree of consistency in our coding process. Some of the decisions we incorporated into the dictionary may have been wrong, but at least they were consistently wrong. After a few more uses and updates, the dictionary may actually be consistently right.

NOTES AND REFERENCES

[1] It should be noted, however, that errors can occur in the industry information obtained from the social security employer file as described by David W. Cartwright in "Major Limitations of CWHS Files and Prospects for Improvement," Policy Analysis with Social Security Research Files, Social Security Administration, Research Report No. 52, 1978, pp. 580-581. See also, Levine, Bruce, "Improving Industry and Place of Work Coding in the Continuous Work History Samples," 1980 American Statistical Association Proceedings, Section on Survey Research Methods.

[2] U.S. Department of Commerce, Bureau of the Census, 1980 Decennial Census Processing Manual, Volume V, Chapter 5, "Industry and Occupation Coding," Section D, "Production Coder Procedures," June 20, 1980.

[3] U.S. Department of Commerce, Bureau of the Census, 1980 Census of Population: Alphabetical Index of Industries and Occupations, Second Edition, May 1981.

[4] U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, Standard Occupational Classification Manual, 1980.

[5] U.S. Department of Labor, Employment and Training Administration, Dictionary of Occupational Titles, Fourth Edition, U.S. Government Printing Office, 1977.

[6] Appel, Martin and Hellerman, Eli, "The Census Bureau Experience with Automated Industry and Occupation Coding," 1983 American Statistical Association Proceedings, Section on Survey Research Methods, pp. 32-40. For experiences with automated or semi-automated occupation coding in other countries, see Lery, Alain and Stephany, Arnaud, "COLIBRI II: Saisie-Chiffrement en Ligne des Bulletins du Recensement de la Population de 1982," Courier des Statistiques, No. 30, April 1984, pp. 20-28; see also Lyberg, Lars, "The Development of Procedures for Industry and Occupation Coding at Statistics Sweden," Statistical Review 1983, No. 5: Essays in Honour of Tore E. Dalenius, pp. 139-156.