CURRENT EARNINGS DIFFERENTIALS OF MEN AND WOMEN: SOME EXPLORATORY REGRESSION ANALYSES

by Joyce A. Stevens and Roger A. Herriot, Bureau of the Census

It is a well-known fact that an earnings differential exists between men and women. Previous studies [1, 3, 4, 5, 6] have attributed this differential to inequalities in male and female occupational distributions, educational attainment, and labor force attachment. What distinguishes this study from others is the inclusion of lifetime work experience as recorded on an individual's Social Security record. One advantage of this data source is that the length of employment, including the number and duration of breaks in "covered" employment, were recorded as they occurred, rather than from personal recall [6]. These employment characteristics were matched to various socio-economic characteristics from the Current Population Survey (CPS), such as occupation, educational attainment and marital status.

This study is a first effort with this new data set to explore the magnitude of wage and salary differentials after controlling for lifetime work experience and other factors related to earnings. Although all income-related variables could not be examined, those[1] used in this study are among the most important and "explain" about 74 percent of the wage and salary distribution for women and 57 percent of that for men. Differences in the distribution of these income-related variables for men and for women are very important for explaining the earnings gap between the sexes. About one-half of the difference in mean wages and salaries of men and women is attributable to the earnings-related variables used in this study. It should be noted that in the analysis which follows, no attempt is made to explain why differences in the distribution of these explanatory variables exist.

Description of Data Set

The data used in this study came from the 1973 Current Population Survey-Summary Earnings Record Exact Match File. This is a public use file prepared as part of a joint Census-Social Security Study[2]. Included on the file are selected social, demographic and work-related variables from the March 1973 Current Population Survey (CPS) which have been matched to longitudinal social security earnings and employment information from the Summary Earnings Record (SER).

The original sample began with over 100,000 records for those individuals interviewed in the March 1973 CPS. Records were eliminated for persons without a social security number, and for persons not between the ages of 25 and 64 or who did not report in the CPS wage and salary income from nongovernment employment of at least one dollar. The final sample includes 21,686 matched persons (13,454 males and 8,232 females).

The results which follow should be interpreted with caution as the sample records were not weighted and because of the biases inherent in the matching process necessary to develop the data set. The biases include mismatches (the bringing together of records which are for different persons) and erroneous nonmatches (the failure to link records which are for the same person). However, the CPS sample is largely self-weighting and the mismatch and nonmatch bias are believed to be small and offsetting[3] hence the unweighted records are probably fairly representative.

Model

Separate earnings functions were estimated for males, for females, and for males and females together. The functional form of the structural equation chosen for this exploratory analysis was

$$(1)\quad E = \gamma B_1^{x_1} B_2^{x_2} \ldots B_{59}^{x_{59}} U$$

where: E = \$ amount of wage and salary earnings.

γ = the expected wage and salary earnings of an assumed reference group.

x_i = a dummy variable (0 or 1) indicating whether a person has a particular characteristic such as 12 years of education or worked full time for 40-47 weeks. (The set of characteristics included in the reference group were, of course, excluded.)

B_i = the proportion by which the expected earnings change as a result of having a particular characteristic rather than the characteristic of the reference group.

U = a random error term.

This particular form has several attractive features. For one thing the expected change in the dollar value of earnings associated with a change in a person's characteristics (x_i) will depend upon the other characteristics of the person. For example, the effect of a change in weeks worked from 30 to 50 will not be a constant amount, but will depend upon the particular occupation which the person has. In addition, one would suspect that variance around alternative expected earning levels due to chance variation would not be a constant amount, but would increase as the expected earnings increases [2].

With a few minor transformations and by taking the log of both sides equation 1 can be put into a form (equation 2) which can be fit by ordinary least squares.

$$(2)\quad \log E = \alpha + b_1 x_1 + b_2 x_2 + \ldots + b_{59} x_{59} + W$$

where: $b_i = \log B_i$

$W = \log U$

If we assume that (U) is log normally distributed, implying that (W) is distributed normally, then the best, linear, unbiased estimates of the coefficients have the additional property of being maximum-likelihood estimators.

For exposition purposes equation (2) can be rewritten as

$$(3)\quad \log E = \alpha + \sum_{i=1}^{9} b_i (SD)_i + \sum_{i=10}^{24} b_i (SE)_i + \sum_{i=25}^{35} b_i (CE)_i + \sum_{i=36}^{59} b_i (LE)_i + W$$

where: SD = the socio-demographic variables including 9 dummy variables pertaining to sex, race, region, SMSA-residence, marital status and presence of children.

SE = the socio-economic variables including 15 dummy variables pertaining to education and occupation.

CE = Current Work Experience including 11 dummy variables pertaining to full-time, part-time status by number of weeks worked.

LE = Lifetime Work Experience including 24 dummy variable classes indicating the number of potential years worked crossed by the number and duration of breaks since age 25.

The advantage of using dummy variables, rather than continuous variables, is that various non-linear or "step effects" of different levels of the variable can be determined. For instance, if education is the independent variable used, the effect on earnings of having 12 years versus 16 years of education can be relatively different from the effect of having 8 years versus 12 years of education. It should be noted that no attempt was made in this study to analyze the effect of interaction effects of various combinations of the independent variables.

Results

The characteristics of the assumed reference or intercept group (α) are used throughout this paper as a basis for comparison. Unless otherwise specified, when a coefficient is discussed, the corresponding characteristic of the intercept group is used as the reference and all other variables are held constant.

The characteristics of persons in the intercept group include:

white, professional, 16 years of education, married, without children present, living in the ring of an SMSA and in the non-South, working full time for 50 to 52 weeks in 1972, and having worked four-quarters of each year for 20 or more contiguous years.

The characteristics included in the intercept group tend to be associated with high earnings;[4] so that deviating from this "norm" reduces one's expected earnings.

In the regression for "both sexes" (not shown in this paper), sex is also entered as an independent variable with "males" in the intercept group. the coefficient of .6488, for females, indicates that on average, women earn 65 percent of what males (who have the same characteristics) earn. However, the average earnings of women in our sample was 31 percent of the average earnings of men (or a 69 percentage point differential to be explained). The difference between the 65 percent and the 31 percent (34 percentage points) can be attributed to differences in work experience, occupation, and the other variables in this study. The 35 percent differential between men and women that has not been accounted for by these variables (100 percent less 65 percent) may be the result of other factors such as job differences within the major occupation groups (i.e., doctor vs. nurse), various discriminatory practices in hiring and promotion, etc. In addition, family responsibilities may deter many women from securing a job for which the pay is commensurate with their marketable skills.

The remaining analysis compares the separate earnings functions (regressions) of males and females. The characteristics of the persons in the two intercept groups are identical to the characteristics of the intercept group in which both sexes were combined, except that "sex" is no longer a variable.

Our focus will be on lifetime work experience, although several other interesting variables will be discussed. Over time men's labor force patterns displayed less discontinuity than women's. Only 15 percent of the women were always 4-quarter workers while 49 percent of the men were. Over half of the women had a break of one or more years in employment compared to 16 percent of the men. The proportion who were at-least-one-quarter workers (a person who had worked at least one quarter, but not four, in a given year) was about the same for men and women (36 percent and 32 percent); however, because generally more men than women work year round, full time it is probable that men in the at-least-one-quarter category worked more quarters over time. This is because over time an "at-least-one-quarter worker" could be (a) a person who worked every quarter each year except for one year in which three quarters were worked or, (b) at the other extreme, a person who only worked 1 quarter each year of continuous employment.

The results suggest that women do not receive the same market returns to continuous work experience as do men.[5] When looking at persons who were always 4-quarter workers, there is some evidence that women with less than 5 years of experience earned 94 percent of what they would have earned had they worked for 20 or more years (the returns for working 5 to 19 years were not significantly different from working 20 or more years), while men with less than 5 years of experience earned 77 percent of what they would have had they worked for 20 or more years (men's earnings increased to 94 percent of the reference group with between 10 and 14 years experience; after 14 years there was no significant difference from working 20 or more years).

Lower monetary returns are attributable to a discontinuous work history (one in which not all of the potential years are worked). We found a consistent pattern of decreasing returns as the coefficients systematically fell from a continuous work history to one with a break in employment of five or more years. For example, when considering 10 to 14 potential years of work experience, men who were always 4-quarter workers had a coefficient of .94, which decreased to .80 for a break in employment of one year, and further fell to .74 for a break of 2 to 4 years. We found that the coefficients decreased more rapidly as the deviations of actual years of work experience from the number of potential years increased. This was true for both men and women. However, due to limitations of the data we were unable to unambiguously analyze the level of earnings associated with particular breaks in employment (including years or quarters not worked). This is because the interpretation of a break in employment of one or more years is hampered by not knowing the number or pattern of quarters worked and because of the multiple meanings for an at-least-one-quarter worker.

Although the lifetime work experience variables had pronounced effects on current earnings for 1972, the variables related to current work experience had the greatest impact on women's and men's earnings. When considering only those categories where a person worked 26 or fewer weeks either full- or part-time, 56 percent of the variation in women's earnings and in excess of 27 percent of the variation in men's earnings were "explained." We found that, relative to year-round, full-time workers of their own sex, women and men had about the same proportional returns for part-time work with the exception of women working 48-52 weeks. These women earned about 8 percentage points more in relative terms than did their male counterparts. Relative to year-round, full-time workers of their own sex, male full-time workers who worked less than 50-52 weeks appeared to receive somewhat higher returns than was the case for comparable women.

The socio-demographic variables used in this study affected the current earnings of men and women differently in many instances. For example, all other things being equal, males of Negro and other races earned 85 percent of what white males earned, while the earnings for females of Negro and other races were not statistically different from those of white females. Relative to being married, being divorced increased women's earnings by 7 percent and decreased men's by 5 percent. While being single, separated or widowed was not statistically different from being married for women, men's earnings were reduced, from 4 percent when widowed to 19 percent when single. Related to this is the finding that the presence of children had no statistical effect on women's earnings but was associated with an increase in men's earnings of 9 percent.

One of the most interesting findings of the study is that the relative return for completing 4 years of college was significantly less for women than for men. There was only a 4.5 percentage point difference between the coefficient for 12 years and 16 years of school for women whereas for men the difference was 22.4 percentage points. Even the 4.5 percent difference is somewhat questionable because of the relatively low t value (1.4)6/ Significant differences in the occupation distribution may help explain this phenomena. Whereas the males were predominantly in occupations which tend to yield higher returns to education (e.g., professional and managerial), we found that slightly over 25 percent of the females were in clerical occupations. One would suspect the "pay-off" for the additional education would be significantly less for these occupations relative to, say professional occupations. If this is true, it could seriously reduce the overall return for women and hence produce the type of results which we found. The question of why such a higher proportion of women with 16 years of education are in clerical occupations will have to await further analysis.

The relative difference in the "pay-off" for education between the sexes narrows for persons when the educational level increased from 16 to 17 or more years. The percentage return for women going on to graduate school relative to four years of college was about twice as great for women as for men (26 versus 13 percent).

Conclusion

After examining the effect of lifetime work experience and other earnings-related variables, there still remains a sizeable unexplained sex differential in earnings. As indicated, further refinement (including interactions) of the variables, expecially lifetime work experience and occupation, would sharpen the analysis. Other variables on the 1973 matched CPS-Social Security data base also might be introduced as predictors, i.e., industry of current employment, income of other family members (particularly spouse), etc.

The data base we have been using here is a rich one for looking at male-female earnings differentials. Perhaps, the exploratory work presented will stimulate others to exploit this publicly available file for furthering their own analyses in this area. (The data file is being distributed by the National Archives Record Service, Machine Readable Archives Division (NNR), Washington, D.C. The file may be purchased from this source under the accession number 375-227.)

Footnotes

*The authors would like to thank Fritz Scheuren (SSA) and Renee Miller, Douglas Sater, and Emmett Spiers (Census Bureau) for their assistance in preparing this paper. Thanks also must be extended to Shirley Roth for the typing.

1/The variables include: lifetime work experience, current work experience, occupation, education, sex, race, region, residence, marital status, and presence of children. Definitions for these variables can be found on Table 1.

2/A more detailed description of the SER File and

factors involved in matching it to the CPS File can be found in, "1973 Current Population Survey-Summary Earnings Record Exact Match File Codebook Reports 4, 5, and 6," from the Studies from Interagency Data Linkages, by Scheuren, et al, Social Security Administration, 1975.

3/Scheuren and Oh, "Fiddling Around with Mismatches and Nonmatches," paper to be published in the Proceedings of the 1975 American Statistical Association Meetings.

4/Although the group of characteristics that will maximize men's earnings will not be identical to those for women, the intercept characteristics chosen for this study approximate a group with very high earnings for both sexes.

5/Similar results were found in,"Statistical Measures of Earnings Differentials," The American Statistician 1975, Vol. 29, No. 1, by J. L. Gastwirth.

6/Even this t-ratio may overstate the true significance of the independent variable. Because of clustering in the CPS sample, the "t" statistics calculated probably tend to overstate the true significance of a coefficient.

References

1. Bergmann, B., "The Economics of Women's Liberation," Annals of the New York Academy of Sciences, Vol. 208, March 1973.

2. Hill, T. P., "An Analysis of the Distribution of Wages and Salaries in Great Britain," Econometrica, Vol. 27, No. 3, July 1959.

3. McNeil, J. and Sater, D., "Recent Changes in Female to Male Earnings Ratios," unpublished paper presented at the Population Association of America meetings in Seattle, Washington, April 1975.

4. Mincer, J. and Polackek, S., "Family Investments in Human Capital: Earnings of Women," Journal of Political Economy, Vol. 82, No. 2, Part II, March/April 1974.

5. Sawhill, I., "The Economics of Discrimination Against Women: Some New Findings," The Journal of Human Resources, Vol. 8, No. 3, Summer 1973.

6. Suter, L., "Occupation, Employment and Lifetime Work Experience of Women," paper presented at the meetings of the American Sociological Association, New York City, August 28, 1973.

Table 1.--Number of Observations and Regression Coefficients for the Independent Variables of the Separate Regressions for Males and Females

VARIABLE NAME	MALES (13,454)			FEMALES (8,232)		
	Number	Coefficient b_i $B_i=10^{b_i}$	t_{b_i}	Number	Coefficient b_i $B_i=10^{b_i}$	t_{b_i}
Intercept Group's Expected Earnings ($\bar{Y}$)	-	$17,685	478.0	-	$ 9,066	204.2
Lifetime Work Experience (1951 to 1972):						
Less than 5 years of Potential Work Experience since age 25:						
Always worked 4 quarters each year	1,332	.7723	-14.6	517	.9352	-1.8
Worked at least 1 quarter, but not 4, each year	433	.6026	-18.8	449	.7732	-6.3
With a break in employment of 1 year	48	.4420	-11.4	99	.6885	-5.7
With one or more breaks in employment totaling 2 to 4 years	11	.8173	- 1.4	50	.6642	-4.7
5 to 9 years of Potential Work Experience since age 25:						
Always worked 4 quarters each year	1,377	.8566	- 8.7	224	1.0462	+0.9
Worked at least 1 quarter, but not 4, each year	668	.7561	-12.5	395	.9024	-2.5
With a break in employment of 1 year	91	.7616	- 5.2	152	.7887	-4.3
With one or more breaks in employment totaling 2 to 4 years	85	.6498	- 7.9	322	.7900	-5.4
With one or more breaks in employment totaling 5 or more years	15	.4075	- 7.1	112	.7188	-5.3
10 to 14 years of Potential Work Experience since age 25:						
Always worked 4 quarters each year	933	.9419	- 3.0	89	1.0382	+0.6
Worked at least 1 quarter, but not 4, each year	537	.8285	- 7.8	195	1.0111	+0.2
With a break in employment of 1 year	101	.7984	- 4.5	89	.8335	-2.7
With one or more breaks in employment totaling 2 to 4 years	110	.7372	- 6.3	285	.9053	-2.2
With one or more breaks in employment totaling 5 or more years	72	.6728	- 6.7	402	.6885	-9.0
15 to 19 years of Potential Work Experience since age 25:						
Always worked 4 quarters each year	706	.9858	- 0.7	71	1.0170	+0.2
Worked at least 1 quarter, but not 4, each year	610	.8692	- 6.1	156	1.0188	+0.3
With a break in employment of 1 year	157	.8626	- 3.6	65	-	-
With one or more breaks in employment totaling 2 to 4 years	144	.8151	- 4.8	167	.8395	-3.3
With one or more breaks in employment totaling 5 or more years	122	.7506	- 6.3	617	.8215	-5.3
20 or more years of Potential Work Experience since age 25:						
Always worked 4 quarters each year*	2,204	(X)	(X)	386	(X)	(X)
Worked at least 1 quarter, but not 4, each year	2,566	.8974	- 7.5	1,421	.8904	-3.6
With a break in employment of 1 year	407	.8416	- 6.5	310	.8461	-3.9
With one or more breaks in employment totaling 2 to 4 years	435	.7881	- 9.2	543	.8726	-3.6
With one or more breaks in employment totaling 5 or more years	290	.7106	-11.0	1,116	.8054	-6.5
Current Work Experience (in 1972):						
Part Time (less than 35 hours per week):						
1 to 13 weeks	62	.0499	-47.6	446	.0481	-98.9
14 to 26 weeks	69	.1699	-29.8	336	.1693	-54.7
27 to 39 weeks	53	.3178	-16.9	230	.3101	-29.1
40 to 47 weeks	32	.3667	-11.6	180	.3540	-23.0
48 to 49 weeks	9	.3357	- 6.7	77	.4180	-12.9
50 to 52 weeks	113	.3839	-20.6	775	.4584	-32.5
Full Time (35 or more hours per week):						
1 to 13 weeks	192	.1458	-53.1	443	.1231	-68.7
14 to 26 weeks	423	.4120	-35.8	577	.3622	-37.5
27 to 39 weeks	612	.6711	-19.3	507	.6193	-17.1
40 to 47 weeks	734	.8185	-10.7	408	.8153	- 6.7
48 to 49 weeks	405	.9175	- 3.5	202	.8610	- 3.5
50 to 52 weeks*	10,752	(X)	(X)	4,251	(X)	(X)
Socio-Economic Variables:						
Education (number of years completed):						
Less than or equal to 8 years	2,374	.6340	-22.9	1,108	.8424	- 4.4
9 to 11 years	2,250	.7143	-17.4	1,551	.8902	- 3.2
12 years	5,063	.7757	-15.2	4,113	.9552	- 1.4
13 to 15 years	1,860	.8283	-10.4	975	.9667	- 0.9
16 years*	1,209	(X)	(X)	360	(X)	(X)
17 or more years	698	1.1264	+ 5.1	125	1.2563	+ 3.8
Occupation (of longest job in 1972):						
Professional, technical and kindred workers	1,594	(X)	(X)	689	(X)	(X)
Managers and administrators	1,888	1.1306	+ 7.2	411	.9294	- 2.0
Sales workers	922	.9618	- 1.9	639	.6976	-10.6
Clerical and kindred workers	754	.8586	- 6.7	2,857	.8320	- 6.9
Craft and kindred workers	3,699	.9721	- 1.7	165	.7980	- 4.3
Operatives, including transport	3,084	.8676	- 8.0	1,676	.7894	- 7.8
Laborers, excluding farm	758	.8352	- 7.5			
Farmers	13	.7011	- 2.6	149	.6183	- 8.7
Farm laborers and supervisors	211	.5702	-14.9			
Private household workers	531	.6921	-13.9	314	.3292	-24.8
Service workers, excluding private household				1,332	.6571	-13.8

See footnotes at end of table.

Table 1.--Number of Observations and Regression Coefficients for the Independent Variables of the Separate Regressions for Males and Females--Continued

VARIABLE NAME	MALES (13,454			FEMALES (8,232)		
	Number	Coefficient $B_i=10^{b_i}$	t_{b_i}	Number	Coefficient $B_i=10^{b_i}$	t_{b_i}
Socio-Demographic Variables (in 1972):						
Race:						
White*.....	12,454	(X)	(X)	7,387	(X)	(X)
Negro and other races.....	1,000	.8480	- 9.7	845	.9927	- 0.3
Region:						
South.....	3,943	.8939	-11.7	2,646	.8983	- 7.5
Non-South*.....	9,511	(X)	(X)	5,586	(X)	(X)
Residence:						
Central City.....	3,660	.9438	- 5.5	2,503	.9734	- 1.7
Ring of an SMSA*.....	5,734	(X)	(X)	3,126	(X)	(X)
Non-SMSA.....	4,060	.8553	-15.1	2,603	.8370	-11.2
Marital Status:						
Married (spouse present)*.....	11,825	(X)	(X)	5,773	(X)	(X)
Single (never married).....	795	.8108	-10.6	669	1.0264	+ 1.0
Separated.....	274	.8960	- 3.6	371	.9815	- 0.5
Divorced.....	433	.9493	- 2.1	748	1.0673	+ 2.3
Widowed.....	127	.9570	- 1.0	671	1.0332	+ 1.3
Presence of Own Children:						
With children.....	8,082	1.0924	+ 8.4	800	.9908	- 0.3
No children present*.....	5,372	(X)	(X)	7,432	(X)	(X)
SUMMARY STATISTICS		MALES			FEMALES	
R^2.....		.56744			.73585	
Adjusted R^2 (R BAR SQUARE).....		.56560			.73408	
F		308.30266			414.11616	

(X) Not applicable.

* Characteristics of the intercept group.

Lifetime Work Experience was calculated after examining a person's earnings history which was covered by Social Security for the period before 1973 and beginning either with 1951 or the year age 25 was attained, whichever came later. A "break" was counted beginning with the first year after 1950 (or the year age 24 was attained, whichever came later) in which the person was not employed, even if the person did not begin employment until after age 25. It should be noted that work experience not covered under the Social Security system, such as employment with the Federal Government, can show up as a "break," thus complicating the interpretation of a "break" which is frequently considered synonymous with "not being employed."

Current Work Experience is a cross of full-time, part-time status for the longest job crossed by the total number of weeks worked in 1972.

Structural Equation: $E=\gamma B_1^{x_1} B_2^{x_2} \ldots B_{59}^{x_{59}} U$

The coefficients (B_i) are interpreted as the proportion by which the expected earnings change as a result of having a particular characteristic rather than the corresponding characteristic of the intercept group. Thus, the coefficient for males of Negro and other races (.8480) can be interpreted as meaning that, ceteris paribus, expected earnings for males of Negro and other races are 84.8 percent of those for white males.

Interpretation of "t" Statistics It should be noted that the CPS is not a simple random sample, and because of clustering in the sample the t-statistics probably overstate the true significance of the independent variables. Thus, particular caution should be exercised when interpreting the B_i term with t-ratios below 2.5.

Source: Derived from the 1973 CPS-SSA Exact Match Study conducted jointly by the Census Bureau and the Social Security Administration.