Volume Title: Distribution of Economic Well-Being

Volume Author/Editor: F. Thomas Juster, ed.

Volume Publisher: NBER

Volume ISBN: 0-884-10478-8

Volume URL: http://www.nber.org/books/just77-1

Publication Date: 1977

Chapter Title: Uncertainty and the Distribution of Earnings

Chapter Author: William R. Johnson

Chapter URL: http://www.nber.org/chapters/c4377

Chapter pages in book: (p. 379 - 396)

10

WILLIAM R. JOHNSON
University of Virginia

Uncertainty and the Distribution of Earnings

LINKS BETWEEN UNCERTAINTY THEORY AND THE DISTRIBUTION OF INCOME

The language of uncertainty theory is familiar to the student of the distribution of income. The same functions which are used by probability theorists to describe outcomes of stochastic events are also used to describe the observed distribution of income or earnings; the normal, the log-normal, and the Pareto distributions are the most familiar examples. It is not necessarily true, however, that the use of these functions to describe the distribution of income implies that income itself is a stochastic variable whose behavior is subject to laws of chance like a roulette wheel. Indeed, many theories of the distribution of income employ functions usually associated with probability densities to describe deterministic models of the distribution of income. Lydall (1968), for example, proposes an entirely deterministic model of income distribution at the upper tail of the distribution which yields the Pareto function. Others have attempted to construct theories which transform symmetrically distributed abilities into the asymmetric, skewed pattern of observed incomes. The hallmark of these theories is that, given an individual's characteristics, his income is entirely determined except possibly for some small error term.

NOTE: The author is a graduate student at the University of Virginia.

While deterministic models of the income distribution borrow the mathematical functions of probability theory, stochastic theories make income itself, to some extent, a chance event or a stochastic variable. The form of the income distribution in these models depends in part on the probability density function of the stochastic elements of income. One such stochastic theory of the income distribution is that of Champernowne (1953), in which income this period depends on income last period in a probabilistic way. Champernowne shows that this Markovian process can ultimately yield a Paretian distribution of incomes, regardless of the initial distribution of income. These results were extended by Mandelbrot (1961).

Milton Friedman has also advanced a stochastic model of income determination (1953). In this case, however, it is not time-dependent stochastic processes which are the key to the model but the year-to-year variation of transitory income around permanent income, which was later to play a central role in Friedman's theory of the consumption function. Friedman was probably overly impressed with the annual variation in incomes as a result of his earlier studies of the income of professionals in independent practice, a group whose incomes probably vary more than the average (Friedman and Kuznets, 1945). The apparent great uncertainty of annual incomes led Friedman to emphasize the fact that the inequality of annual incomes is greater than the inequality of permanent lifetime incomes and, by extension, the distribution of utility. Friedman went farther than this in his discussion of uncertainty; he brought out the role of risk in the choice of occupation. That is, some individuals choose occupations in which the dispersion of incomes is greater than in other occupations. Chance would govern their permanent incomes as well as annual incomes. However, by emphasizing the role of choice as well as chance in the process (since an individual selects among different lines of work), Friedman again cautioned against interpreting income inequality as reflecting inequality of, in this case, expected utility. In part, the distribution of income is unequal because some individuals make risky choices and earn income with varying degrees of success.

Recent empirical studies are beginning to echo Friedman's appeal to uncertainty in explaining income inequality. Jencks (1972), especially, finds that none of the familiar determinants of earnings—demographic characteristics, family background, schooling, cognitive ability—explain more than a small fraction of the observed inequality of income in the United States. Jencks ascribes some of the unexplained variance to random factors: "In general, we think that luck has far more influence on income than successful people admit" (1972, p. 227). Although Jencks's results are open to serious question because of his peculiar model and the heterogeneity of his data sources, his inability to explain much of the dispersion in earnings has been duplicated by other studies. Recent work using microdata files has failed to explain more than a small fraction of the total variance in individual earnings. For example, Taubman and Wales's recent work with the extremely rich Thorndike-Hagan file of veterans emphasizes the failure of deterministic models to explain much income inequality (Taubman and Wales, 1973). Their earnings function, which included not only traditional demographic variables, but also extensive measures of ability taken from armed forces test results, and family background variables, had R^2's of around .10. Recent work by Paul Taubman (reported in this volume) extends the Taubman-Wales results and raises the explained variance of earnings to more than 40 percent. Hall (1973) estimated wage equations for race-sex groups from Survey of Economic Opportunity data and found standard errors of estimate for his equations of nearly forty cents per hour. These results indicate that earnings are hard to predict; there seems to be a good deal of indeterminacy in earnings.

Taubman and Wales performed an experiment which sheds light on the structure of uncertainty in earnings. In their estimate of 1969 earnings, they used as an explanatory variable the error from their equation estimating earnings in 1955 for the individual. In this way, any permanent, yet unobserved, explanation of earnings in 1955 would help to explain earnings in 1969. Interestingly, the inclusion of these residuals raised the R^2 from .10 to .33, indicating that these permanent unobserved factors account for close to one-quarter of the variance in incomes (Taubman and Wales, 1973, p. 37). As for the rest of the variance in incomes, Taubman and Wales concluded that "... two-thirds of the variation in earnings in any year represents either random events such as luck, and/or changes in underlying characteristics" (1973, p. 38).

TYPES OF UNCERTAINTY IN INCOMES

Uncertainty not only plays a role in several theories of the distribution of income, but it appears that empirical estimates confirm the existence of a large amount of uncertainty or unexplained variance in the estimation of earnings. To the extent that uncertainty plays a role in the determination of earnings, the probability density functions of the stochastic terms of the earnings function determine, in part, the shape of the distribution of earnings. The stochastic elements in earnings are of two basic types. The first is the year-to-year variability in earnings which led Friedman to distinguish between observed income and permanent income. The importance of this short-run variability has been discounted by Thurow who says: "The distribution of lifetime incomes probably looks very similar to the distribution of annual incomes" (1969, p. 108). There is, however, another facet of uncertainty which transcends the year-to-year variations—unexplained permanent dispersion in incomes. The permanent uncertainty is clearly one key to the inequality of lifetime incomes among individuals and is, therefore, the focus of this paper.

CAUSES OF UNEXPLAINED PERMANENT DISPERSION IN EARNINGS

What accounts for this unexplained permanent variation in earnings? Certainly, some of the explanation rests with actual differences in individuals which are not, or cannot be, observed by the statistician. Although the Taubman and Wales data encompassed a broad range of explanatory variables, there are undoubtedly unmeasured differences in personality, motivation, and ability which create real differences in individual productivities. In this case, apparent uncertainty in incomes is due not to stochastic variables in the earnings function but to unobserved differences in characteristics. That is, firms pay wages which reflect marginal products and recognize differences in productivity among individuals; however, outside observers cannot recognize or measure all of the factors which affect an individual's productivity. These unobserved components clearly look like earnings uncertainty to the outside observer and they may be just as uncertain a priori to the individual himself. A person may not be aware, in advance, of all of the factors which affect his productivity and, to this extent, unobserved components are uncertain to him, too.

Another source of differences in earnings of observationally equivalent persons are compensating wage differentials for nonmonetary aspects of jobs. In other words, part of earnings may be payments for job characteristics rather than for personal characteristics. Lucas in an empirical investigation of job characteristics finds rather perverse results; many characteristics of jobs which are usually considered to be unpleasant are rewarded negatively rather than positively ceteris paribus (Thurow and Lucas, 1972). Greg Duncan (in this volume) studies the impact of fringe benefits in reducing wage differentials.

A third reason for unexplained differences in earnings is one which has been emphasized recently in another context—costly information and imperfect markets. This rubric embraces two principal hindrances to perfectly competitive labor markets—first, the institutional forces of unions and industrial structure which prevent the achievement of perfect competition; and second, the costs of acquiring information both for the worker and for his (potential) employer, and the related costs of mobility between occupations and geographic locations. To the extent that information is imperfect and mobility is costly, then differences in earnings of persons with exactly the same productivity will not be eliminated. An estimated earnings function would not be able to explain all the variance in earnings even if we could measure every characteristic which affects an individual's productivity.

THIS PAPER

The dispersion of permanent lifetime incomes among observationally equivalent persons can be ascribed either to unobserved differences in personal characteristics which affect an individual's productivity or to imperfections in labor markets which allow equally productive individuals to be paid different amounts. The subject of this paper is not the cause of uncertainty in earnings but rather the effect of uncertainty on the wage structure itself. The principal focus is on differences in uncertainty across occupations: Is there a tradeoff between risk and return in occupational earnings? The importance of this question as regards the distribution of income is at least twofold. First, if average earnings are higher in occupations which have greater unexplained variance in earnings (and therefore less certain earnings to the potential entrant to the occupation), then the market is compensating for uncertainty, and the distribution of expected utility is likely to be less unequal than the distribution of observed incomes. Second, the positive correlation of dispersion with average earnings provides an explanation for the observed skewness in the distribution of labor incomes.

In the rest of this paper, the tradeoff between risk and return in the labor market is estimated, using data from the 1970 Census. In order to estimate uncertainty, an earnings function is specified. Because many significant variables cannot be observed, estimates of uncertainty are necessarily biased. The effect of this bias on the results is considered. Results for different demographic groups indicate that the risk-return relation exists and seems to be related to the degree of occupational immobility in the demographic group.

A SIMPLE MODEL OF EARNINGS

Consider a simple linear model of permanent, lifetime earnings. Intertemporal variation of earnings is not important here, only the dispersion of earnings across individuals. Let

$$Y_{ij} = \alpha_i + \beta_i Z_j + \delta_i A_j + \tilde{X}_{ij} \tag{1}$$

where

Y_{ij} is the permanent earnings of individual j in occupation i;
α_i is a parameter (scalar);
Z_j is a vector of observed characteristics of individual j;
A_j is a vector of unobserved characteristics of individual j;
β_i, δ_i are parameter vectors conformable to Z_j, A_j; and
$\tilde{X}_{ij}$ is the stochastic element of earnings for individual j in occupation i.

Assume that $\tilde{X}_{ij}$ is distributed identically for all individuals in occupation i, although clearly the value $\tilde{X}_{ij}$ takes on may be different for each person. The expected value of $\tilde{X}_{ij}$ is zero. An individual's earnings in occupation i are determined by his characteristics, Z_j and A_j, the reward occupation i puts on these characteristics, and a chance element, $\tilde{X}_{ij}$. This stochastic term is permanent for the individual—his one draw from the distribution determines his earnings in the occupation forever.

In an econometric earnings function, both A_j and $\tilde{X}_{ij}$ are pushed into the error term of the equation. They must be separated, at least conceptually, because the unobserved characteristics, A_j, may be known by the individual choosing an occupation, whereas $\tilde{X}_{ij}$ cannot be known, by definition. Inability to observe A_j will introduce bias into the estimates of uncertainty; unexplained variance in earnings for individuals with the same Z_j could arise from either $\tilde{X}_{ij}$ or differences in A_j.

Given the earnings function of (1), an individual is assumed to maximize his lifetime expected utility by choosing the appropriate occupation. For an individual with a given set of attributes, Z_j and A_j, the choice is made on the basis of the α_i, β_i, δ_i, and the distribution of $\tilde{X}_{ij}$ for each occupation. Clearly, no person actually knows α_i, β_i, and δ_i; what they may know is mean earnings in occupation i for individuals with similar characteristics, and an impression of the dispersion of earnings around the mean. In fact, Freeman (1971) finds that college students are quite well informed about the pattern of earnings in various occupations. By maximizing expected utility in choosing an occupation, persons will be basing their choice both on mean earnings and the uncertainty, or dispersion, of earnings. Analogous to financial theory, equilibrium in the labor market should involve a tradeoff between risk and return across occupations for given demographic groups. Occupations with a great deal of uncertainty must offer greater than average wages in order to entice entrants. One of the key assumptions here is the lack of mobility between occupations. To the extent that individuals move between occupations, the risk-return relationship is mitigated; a person who fares poorly in one occupation will move to another. In reality, given the costs of mobility (including the training costs for the new occupation), there is an optimal pattern of occupational "search," similar to the job search process described by Mortensen (1970). It suffices to assert that the more mobility between occupations, the weaker the relationship between risk and return.

THE DATA

Earnings distributions for full-time workers broken down by race, age, sex, education and occupation are given in the 1970 Census. Each distribution gives the dispersion of full-time earnings for individuals in a given age, sex, race, and education group for a particular occupation. By comparing mean earnings and dispersion across occupations, *within* a given demographic group, the contribution of Z_i to earnings does not have to be explicitly estimated. For the time being, we assume that all individuals in a particular demographic group (same Z's) have the same unobserved characteristics, A_i. I shall later investigate the bias introduced into our estimates by such an assumption.

FUNCTIONAL FORM

In order to specify a functional form for estimation, both a utility function and a distribution function for $\tilde{X}_{ij}$ must be postulated. The traditional use of mean and variance is not necessarily appropriate; recent work in the theory of uncertainty shows that the mean-variance criteria can be inconsistent with the expected utility hypothesis (see Feldstein [1969]). A convenient hypothesis is that earnings and the stochastic term follow a log-normal distribution, whereas the utility function is of the constant relative risk aversion type. Then, expected utility can be expressed as a linear function of the log of the mean of income and the variance (see Weiss [1972]). On the other hand, if earnings are normally distributed, then utility can be written as a simple function of mean and variance. In this study, many different functional forms were tried.

Since the distribution data come grouped into rather large income classes, the data were fitted to both the normal and the log-normal distributions in order to estimate parameters of the earnings distributions.[1] Otherwise, the error in estimating variances directly from grouped data would be large, especially with the open-ended highest income bracket. Then, the estimated parameters of these normal and log-normal approximations were used to estimate risk-return relationships. The dependent variable in each estimated equation was either the mean earnings in the occupation for the particular demographic group, as computed by the Census Bureau, or the log of that mean.

RESULTS

The results cannot be easily summarized. No one specification was clearly superior for all 36 demographic groups (2 race groups, 3 age groups and 6 education groups). Neither the normal nor the log-normal approximation to the distribution of earnings was clearly superior for all demographic groups, by a chi-squared test of goodness of fit, so both were used. For every demographic group, the estimated coefficient on the dispersion variable, the standard deviation of earnings, was positive. Even when the coefficient of variation was used as the dispersion variable, the results were positive when earnings were fitted to the log-normal distribution. As an example of the results, Table 1 gives the results for all demographic groups for one of the estimated equations. Results for blacks were less definite because earnings distributions were available for fewer occupations.

Because the estimated coefficient on the dispersion variable was positive in virtually every case, it appears that one can say with a fair degree of assurance that uncertainty is compensated to some degree in the labor market. However, at least two caveats should be appended to this assertion. First, the measure of uncertainty used, dispersion of 1969 full-time earnings, does not necessarily represent the permanent dispersion of earnings. To the extent that annual earnings fluctuate around permanent earnings, this measure of dispersion overstates the permanent dispersion in earnings. Only a longitudinal sample could reveal the extent of transitory variation in incomes. Friedman and Kuznets's data for independent professional incomes are probably not representative (Friedman and Kuznets, 1945). A second qualification to the results is that the risk of unemployment seems to exert a negative influence on relative occupational earnings in many of the samples. This influence is measured by the variable FULLEMP, which is high when the risk of unemployment or part-time work is low. Thus, the market does *not* seem to compensate for the risk of unemployment undertaken in a given occupation, a relationship Hall (1970) has suggested for geographic areas.[2] Because we have used cross-sectional data from a generally low unemployment year (1969), we have certainly not measured adequately the effect of *cyclical* unemployment on average occupational earnings.

TABLE 1 Regression Results: Effect of Earnings Dispersion on Mean Earnings across Occupations for Various Demographic Groups (Dependent variable: mean occupational earnings; *t*-ratios in parentheses)

		Regression Coefficients				
Educ.	N	FULLEMP[a]	GRWTH[b]	Standard Deviation[c]	Constant	R^2 Adj.
		Male Whites, Aged 25–34				
0–8	55	6,299	9.856	.7141	−1,016	.254
		(1.344)	(2.003)	(3.397)	(.2233)	
9–11	84	3,100	8.348	.3708	4,163	.178
		(1.364)	(3.007)	(3.007)	(1.843)	
12	101	−1,633	8.625	1.048	7,110	.245
		(.470)	(3.734)	(4.365)	(1.962)	
13–15	100	18,357	7.728	.6269	−10,273	.400
		(6.985)	(3.184)	(3.768)	(3.902)	
16	90	11,829	7.409	.5332	−2,695	.133
		(2.863)	(1.856)	(1.943)	(.6864)	
17+	73	201	3.908	1.111	6,733	.156
		(2.857)	(.8189)	(3.847)	(4.989)	
		Male Whites, Aged 35–54				
0–8	83	8,601	12.05	.6224	−1,945	.293
		(2.893)	(3.076)	(3.151)	(.6711)	
9–11	96	14,792	12.969	1.144	−8,590	.309
		(2.361)	(3.117)	(4.631)	(1.3471)	
12	107	14,523	14.95	1.689	−9,970	.535
		(3.158)	(4.507)	(8.261)	(2.267)	
13–15	103	28,681	14.108	1.837	−23,722	.708
		(5.317)	(4.643)	(12.04)	(4.606)	
16	94	331.5	11.085	1.923	4,086	.573
		(5.075)	(2.489)	(10.615)	(4.450)	
17+	79	569.2	4.856	2.044	4,452	.533
		(7.322)	(.661)	(8.814)	(3.132)	
		Male Whites, Aged 55–65				
0–8	77	1,634	20.07	.3981	5,268	.238
		(.8418)	(4.278)	(1.580)	(2.662)	
9–11	83	3,344	21.79	.4137	4,236	.268
		(1.691)	(4.672)	(1.990)	(2.073)	
12	95	283.6	25.55	1.462	4,104	.422
		(3.656)	(4.206)	(7.217)	(4.705)	

TABLE 1 (continued)

		Regression Coefficients				
Educ.	N	FULLEMP[a]	GRWTH[b]	Standard Deviation[c]	Constant	R^2 Adj.
		Male Whites, Aged 55–65 (continued)				
13–15	86	18,509	18.93	1.110	−11,713	.482
		(3.880)	(3.829)	(5.734)	(2.518)	
16	71	264	20.74	1.706	5,391	.378
		(4.123)	(2.950)	(6.261)	(3.690)	
17+	56	35,984	10.951	.861	−22,858	.250
		(2.106)	(1.044)	(3.043)	(1.431)	
		Male Blacks, Aged 25–34				
0–8	19	−31.71	15.50	.2187	4,403	.589
		(1.3203)	(2.828)	(2.196)	(13.55)	
9–11	24	.748	18.795	.511	4,348	.727
		(.0211)	(6.345)	(1.950)	(5.609)	
12	30	7,174	7.120	.755	−2,267	.383
		(1.532)	(1.857)	(2.377)	(9.518)	
13–15	23	6,867	4.671	−.148	1,517	–
		(.986)	(.922)	(.2753)	(.2443)	
16	11	−6,818	1.949	.952	12,294	–
		(.798)	(.1332)	(.9377)	(1.692)	
17+	6	23,333	−12.30	−.1138	−9,508	–
		(.728)	(.423)	(.051)	(.2704)	
		Male Blacks, Aged 35–54				
0–8	25	5,665	14.84	.297	−890	.288
		(1.232)	(2.609)	(1.549)	(.207)	
9–11	28	8,975	17.36	−.007	−2,006	.322
		(1.687)	(2.504)	(.0369)	(.422)	
12	34	26,451	11.428	.1225	−18,254	.558
		(4.385)	(2.204)	(.8717)	(3.165)	
13–15	26	28.23	7.184	.1208	7,533	–
		(.205)	(1.105)	(.340)	(5.883)	
16	15	278.9	9.179	1.372	3,939	.615
		(4.952)	(.945)	(3.237)	(2.448)	
17+	10	40,331	−2.363	2.512	37,565	–
		(2.231)	(.2337)	(6.728)	(2.393)	
		Male Blacks, Aged 55–64				
0–8	19	6,037	6.696	.5212	−1,873	.463
		(1.828)	(1.075)	(2.967)	(.602)	

TABLE 1 (concluded)

		Regression Coefficients				
Educ.	N	FULLEMP[a]	GRWTH[b]	Standard Deviation[c]	Constant	R^2 Adj.
9–11	20	−6,824	6.125	.908	9,981	.735
		(1.828)	(1.075)	(2.967)	(.602)	
12	15	−2,814	−3.889	.560	8,180	-
		(.373)	(.493)	(1.590)	(1.110)	
13–15	8	2,840	−.306	1.745	−842	.458
		(.193)	(.030)	(2.560)	(.054)	

SOURCE: 1970 Census of Population PC(2)-8B, *Earnings by Occupation and Education*.
[a]FULLEMP is the ratio of average earnings of all workers to average earnings of full-year workers. A high figure indicates little part-time work or unemployment.
[b]GRWTH is the rate of growth of the occupation from 1960 to 1970.
[c]Standard deviation is the standard deviation of earnings, under the normal approximation.

An indication of the economic significance of the results displayed in Table 1 is given in Table 2, which presents the dollar impact on mean wages of a change in the dispersion variable of one standard deviation. Clearly, not only are the results statistically significant, but they are also economically significant, with dollar effects ranging from $277 to $4,124.

UNCERTAINTY AND IMMOBILITY

As mentioned above, the risk-return tradeoff should have a steeper slope, the less mobility there is between occupations. When persons are locked into their choices of occupations, then there is no opportunity for an ex post equalizing movement between occupations. Dispersion in earnings can exist under such circumstances. If, on the other hand, mobility is perfect between occupations, then there should be no differences in earnings among occupations which do not stem from differences in individual characteristics or nonpecuniary factors. Another way to see this is to consider that with perfect mobility, there is no penalty for choosing a highly risky occupation, because one can leave costlessly if one's luck is bad.

Interestingly, the data do support a tentative conclusion that immobility may be related to the strength of the tradeoff between risk and return. Figure 1 presents data on mobility between broad occupational classes—broader, in fact, than the units of observation for the risk-return equations. The data are grouped to correspond most closely with the

TABLE 2 Dollar Effects of Mean Occupational Earnings of a One Standard Deviation Change in the Dispersion Variable: Equation Reported in Table 1, White Males

Education	Effect on Mean Occupational Earnings
	Age 25–34
0–8	$ 519
9–11	277
12	567
13–15	443
16	423
17+	1,208
	Age 35–54
0–8	544
9–11	927
12	1,480
13–15	1,974
16	2,722
17+	4,124
	Age 55–65
0–8	327
9–11	414
12	2,184
13–15	1,373
16	3,106
17+	2,029

groupings of our estimated equations. As expected, immobility rises sharply from the 20–34 to the 35–49 age group and slightly beyond that. Immobility also rises with education in the two older age groups, while it tends to fall with education in the lowest age group. It is well to caution that these results may be peculiar because of the breadth of occupational classes involved in the definition of mobility.

In Figure 2, the estimated coefficients on the dispersion term (standard deviation of earnings) are plotted for the equation whose complete results appear in Table 1. In fact, the coefficients seem to behave quite like the measures of immobility in Figure 1. For the middle age group, effect of dispersion increases with education; for the younger age group,

FIGURE 1 Immobility: Proportion of Individuals in 1970 with Same Major Occupation as 1965; by Education and Age, Males

SOURCE: 1970 Census.

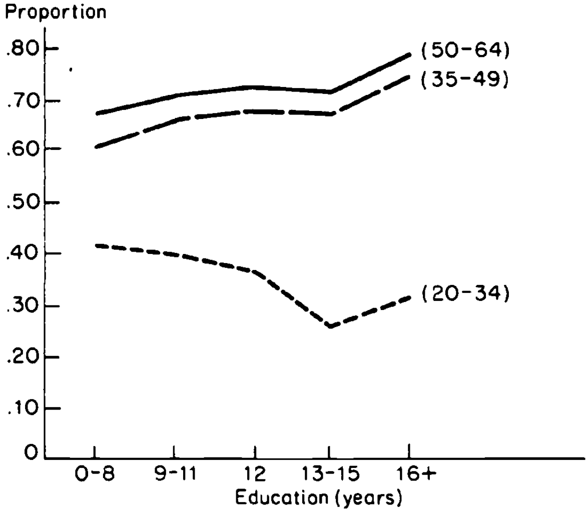
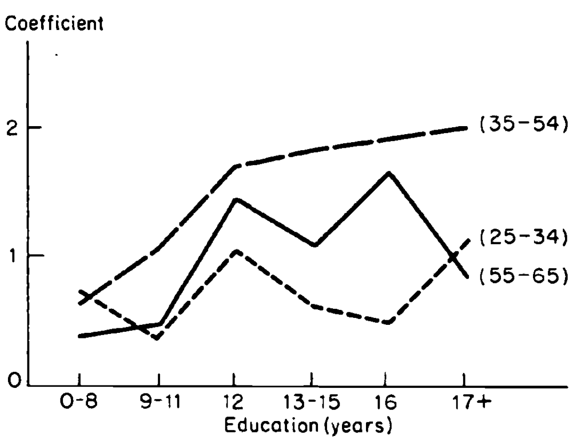


FIGURE 2 Regression Coefficients: Dispersion Term (Standard Deviation, Normal Approximation) by Education and Age; White Males

SOURCE: Table 1.



there is not such a clear pattern. While it is probably incorrect to make too much of this similarity, the evidence does point both to a positive relationship between risk and return in the labor market and to an association of the goodness of fit or strength of this relationship and occupational immobility. The robustness of these conclusions is underscored by the large number of demographic groups and functional forms used in making the estimates.

THE PROBLEM OF OMITTED VARIABLES

So far, the observed association between the dispersion of earnings and mean earnings across occupations, but within demographic groups, has been attributed to the tradeoff between risk and return in the labor market. In this section of the paper, another explanation for these results is considered—the omission of ability as a variable.

The bias introduced by the omission of A_j or ability from the estimates causes two kinds of error: first, the measured dispersion of the random term, $\tilde{X}_{ij}$, for an occupation includes the dispersion of A_j in addition to the true dispersion of $\tilde{X}_{ij}$; second, the effect of differences in mean ability across occupations is omitted from the equation explaining differences in mean earnings across occupations. In the mean-variance framework (to make the analysis easier), assume that the "true" relationship for a given demographic group is

(2) $$Y_i = K + \gamma(\text{var } \tilde{X})_i + \beta A_i$$

or

$$\hat{Y}_i = Y_i - \beta A_i = K + \gamma(\text{var } \tilde{X})_i$$

where

Y_i is mean earnings in occupation i;

$(\text{var } \tilde{X})_i$ is the variance of "true" $\tilde{X}$ in occupation i;

A_i is the mean ability of individuals in occupation i; and

K, γ, and β are parameters.

The coefficient γ is the payment for risk which we are trying to estimate. Note that we initially assume that occupations reward ability identically, that β is the same for all i.

Instead of estimating this "true" relation, we estimate

(3) $$Y_i = k + g[\text{est } (\text{var } \tilde{X})_i]$$

where est $(\text{var } \tilde{X})_i$ is the measured variance of earnings within occupation

i. The purpose of this analysis is to discover the extent to which the estimated parameters, k and g, differ from the true parameters, K and γ.

The measured variance of earnings, est (var $\tilde{X}$)$_i$, can be written

$$\text{(4)} \qquad \text{est}\,(\text{var}\,\tilde{X})_i = (\text{var}\ Y)_i = (\text{var}\,\tilde{X})_i + \beta^2(\text{var}\,A)_i$$

where (var A)$_i$ is the variance of A within occupation i and $\tilde{X}_{ij}$ is independent of A. Considering only large-sample properties of the estimate,

$$\text{plim}\, g = \frac{\text{cov}\,[(Y_i), \text{est}\,(\text{var}\,\tilde{X})_i]}{(\text{var}\ Y_i)}$$

$$= \frac{\text{cov}\,[(\hat{Y}_i + \beta A_i), (\text{var}\,\tilde{X})_i + \beta^2(\text{var}\,A)_i]}{(\text{var}\ \hat{Y}_i) + \beta^2(\text{var}\,A_i) + \beta\ \text{cov}\,(\hat{Y}_i, A_i)}$$

Thus

$$\text{(5)} \qquad \text{plim}\, g = \gamma \frac{\text{var}\ \hat{Y}_i}{\text{var}\ \hat{Y}_i + \beta^2(\text{var}\,A_i)} + \frac{\beta\ \text{cov}\,[A_i, (\text{var}\,\tilde{X})_i] + \beta^2\ \text{cov}\,[\hat{Y}_i, (\text{var}\,A)_i] + \beta^3\ \text{cov}\,[A_i, (\text{var}\,A)_i]}{(\text{var}\ \hat{Y}_i) + \beta^2(\text{var}\,A_i)}$$

If abilities are randomly distributed among occupations, then all the covariance terms will be zero in the probability limit, and var A_i, the variance of average ability across occupations, is zero. In this case, there is no bias because plim $g = \gamma$. In fact, these are reasonable assumptions, given the assumptions of the model (there is no reason to expect any particular pattern of abilities across occupations if β is the same for each occupation), so that in the probability limit, abilities should be identically distributed across occupations.

If, on the other hand, β differs from occupation to occupation, then (5) becomes

$$\text{(6)} \qquad \text{plim}\, g = \gamma \frac{\text{var}\ \hat{Y}_i}{\text{var}\ \hat{Y}_i + \text{var}\ \beta_i A_i} + \frac{\text{cov}\,[\beta_i A_i, (\text{var}\,\tilde{X})_i] + \text{cov}\,[\hat{Y}_i, \beta_i^2(\text{var}\,A)_i] + \text{cov}\,[\beta_i A_i, \beta_i^2(\text{var}\,A)_i]}{\text{var}\ \hat{Y}_i + \text{var}\ \beta_i A_i}$$

Again, the first and second covariance terms are zero in the probability limit, but the third will be positive even if A_i and (var A)$_i$ are not correlated because β_i is positive. Thus, (6) becomes

$$\text{(7)} \qquad \text{plim}\, g = \gamma \frac{\text{var}\ \hat{Y}_i}{\text{var}\ \hat{Y}_i + \text{var}\ \beta_i A_i} + \frac{\text{cov}\,[\beta_i A_i, \beta_i^2(\text{var}\,A)_i]}{\text{var}\ \hat{Y}_i + \text{var}\ \beta_i A_i}$$

If individuals do not know their abilities and are therefore choosing occupations regardless of β_i, then there may again be random assortment of abilities to occupations. In this case, A_i and $(\text{var}\,A)_i$ will again be constant, in the probability limit, across occupations. Thus, var $\beta_i A_i$ will equal A_i var $\beta_i > 0$ and $\text{cov}[\beta_i A_i, \beta_i^2 \text{ var}(A)_i]$ will depend on $\text{cov}(\beta_i, \beta_i^2)$. The net distortionary effect will be indeterminate because $\text{var}(\beta_i A_i)$ decreases g while $\text{cov}[\beta_i A_i, \beta_i^2(\text{var}\,A)_i]$ increases the estimate.

Curiously, if individuals do know their abilities and β_i, the distortion becomes less positive. In this case, high ability individuals will choose occupations which reward ability highly, so that $\text{cov}(\beta_i, A_i) > 0$, and therefore, $\text{var}(\beta_i A_i) > A_i \text{ var}(\beta_i) > 0$. The denominator of the first term of (7) is larger and the bias becomes less positive. In fact, if the covariance term in (7) were zero, then the estimate g would underestimate γ. In fact, however, $\text{cov}[\beta_i A_i, \beta_i^2(\text{var}\,A)_i]$ will be positive even if A_i and $(\text{var}\,A)_i$ are constant over i, because β_i is positively correlated with β_i^2.

Although there is no evidence on β_i, there is some evidence that $\text{cov}[A_i, (\text{var}\,A)_i]$ may be negative. Data gathered during World War II which matched civilian occupations of military men with scores on armed forces tests show that, as expected, people in higher status or higher income occupations tend to perform better on tests (Harrell and Harrell, 1945). However, the variance of ability within an occupation, as measured by test scores, tends to decrease as the average ability of the occupation increases. This result may be explained by the assumption of a minimum level of ability which differs from occupation to occupation. Occupations with higher standards will have higher average abilities and, because a smaller range of abilities is acceptable, a smaller ability variance. Under these circumstances, $\text{cov}[\beta_i A_i, \beta_i^2(\text{var}\,A)_i]$ will be smaller than it would be if A_i and $(\text{var}\,A)_i$ were the same for all i, but still may be either positive or negative. Thus, one cannot say for sure whether it will be overestimated or underestimated, but at least there is no definite positive bias.

CONCLUSION

In this paper, a stochastic model of earnings was used to derive estimable relations between risk and return in the labor market. The estimates reveal a systematic positive effect of earnings dispersion on average earnings. This result holds up for many different demographic groups, and for both the normal and log-normal approximations to the distribution of earnings. Furthermore, the estimated coefficients for the dispersion term are larger for demographic groups which are less mobile between occupations, conforming to theoretical expectations. Finally, the omission of certain key variables, while a problem, may not be a serious source of bias in the estimates.

There are two implications of these findings for the distribution of income. First, since the market compensates for risk, in the form of earnings uncertainty, the distribution of expected utility is less unequal than the distribution of incomes. However, the reverse is true if the risk of unemployment is considered. Second, if the variance of earnings around an occupational mean is positively related to the occupational mean, then there is a "real" explanation for the rightward skewness in the distribution of earnings. Consider the third moment of the distribution, a traditional measure of skewness

$$S = \sum_i \sum_j (Y_{ij} - \bar{Y})^3$$

where, as before, i indexes occupations, j indexes individuals. If earnings within an occupation are not skewed around the occupational mean, and the means themselves are not skewed around the population mean, then

$$S = 2 \sum_i (\text{var } Y)_i (\bar{Y}_i - \bar{Y})$$

where $(\text{var } Y)_i$ is the variance of earnings within occupation i and $\bar{Y}$ is the population mean. Clearly, the size of S depends on the positive correlation between $(\text{var } Y)_i$ and $\bar{Y}_i$.

NOTES

1. The normal and log-normal approximations were based on a procedure in Aitchison and Brown (1957).
2. Another variable, GRWTH, the percentage change in employment in the occupation, 1960–70, was a proxy for changing demand. As expected, its coefficients were positive and significant. The age-earnings profiles were tried as variables but were not significant.

REFERENCES

Aitchison, J., and Brown, J. A. C. *The Lognormal Distribution.* Cambridge: Cambridge University Press, 1957.

Champernowne, D. G. "A Model of Income Distribution," *Economic Journal* 68 (June 1953).

Feldstein, M. "Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection." *Review of Economic Studies* 36 (Jan. 1969).

Freeman, R. *The Market for College-Trained Manpower.* Cambridge, Mass.: Harvard University Press, 1971.

Friedman, M. "Choice, Chance, and the Personal Distribution of Earnings." *Journal of Political Economy* 61 (Aug. 1953).

———. *A Theory of the Consumption Function.* Princeton: Princeton University Press for NBER, 1957.

Friedman, M., and Kuznets, Simon. *Income from Independent Professional Practice.* New York: NBER, 1945.

Hall, R. "Why Is the Unemployment Rate so High at Full Employment?" *Brookings Papers on Economic Activity*, 3: 1970.

———. "Wages, Income and Hours of Work in the U.S. Labor Force." In Glen Cain and Harold Watts, eds., *Income Maintenance and Labor Supply.* New York: Academic Press, 1973.

Harrell, T., and Harrell, M. "Army General Classification Test Scores for Civilian Occupations." *Educational and Psychological Measurements* 5 (Autumn 1945).

Jencks, C. *Inequality.* New York: Basic Books, 1972.

Lydall, H. *The Structure of Earnings.* Oxford: Clarendon Press, 1968.

Mandelbrot, B. "Stable Paretian Random Functions and the Multiplicative Variation of Income." *Econometrica* 29 (Oct. 1961).

Mortensen, D. "Job Search, the Duration of Unemployment, and the Phillips Curve." *American Economic Review* 60 (Dec. 1970).

Taubman, P., and Wales, T. "Higher Education, Mental Ability and Screening." *Journal of Political Economy* 81 (Jan./Feb. 1973).

Thurow, L. *Poverty and Discrimination.* Washington, D.C.: Brookings Institution, 1969.

Thurow, L., and Lucas, R. E. B. *The American Distribution of Income: A Structural Problem.* Joint Economic Committee, Mar. 1972.

Weiss, Y. "The Risk Element in Occupational and Educational Choices." *Journal of Political Economy* 80 (Nov./Dec. 1972).