THE RATE OF CHANGE IN THE SIZE DISTRIBUTION OF WAGES AS A VECTOR

by Henry J. Cassidy*

Planning Research Corporation and George Washington University

This paper attempts to measure the rate of change in the size distribution of wages over time in a rigorous, analytic way, and to relate that change to the business cycle. The basic problem for which this paper provides a solution is to relate changes in a size distribution to levels of and changes in single-dimensioned variables (unemployment, Gross National Product, and the consumers price index). Let F stand for the cumulative relative size distribution of wages, a function of wages. F takes on values zero through one. Let $\bar{F}$ be a given value of F, e.g., $\bar{F} = 0.25$. The proposed solution to the basic problem is to measure the rate of change in consecutive F's at $\bar{F}$. The composite of such measurements at $\bar{F}$ over time forms a vector, the length of which depends upon the number of time periods observed. The number of vectors thus derived depends upon the number of values of $\bar{F}$ selected. The various vectors are then related to the general economic conditions and the respective values of $\bar{F}$. The general economic conditions have a differential effect on the various vectors; e.g., those wage earners with relatively low wages are affected differently by a given turn of the business cycle than are those with high wages.

The paper includes several supplementary investigations: (a) estimating each of the annual cumulative relative size distributions of wages for a specific analytic function, (b) relating analytically the size distribution construct to the Lorenz curve concept and the Gini coefficient, (c) predicting and simulating size distributions for various economic conditions, (d) formulating tax trade-offs, and (e) suggesting further uses and extensions.

Economists have analyzed size distributions of incomes in many ways, attempting to answer the following general questions.[1] What analytic form underlies the distributions? Why is a particular analytic form appropriate? What can size distributions say about the inequality of incomes? How are the size of incomes and the inequality of incomes related to socio-economic variables?[2]

The question this paper attempts to answer is somewhat different: how does the entire size distribution of wages[3] change as economic conditions change, and, consequently, what is the effect on wage inequality? The question has been raised by others, but not answered. Champernowne, for example, briefly mentions but does not take up the possibility of analyzing changes in consecutive distributions "corresponding to various economic situations in order to deduce . . . the

*This study is an extension of several chapters of my unpublished Ph.D. Dissertation, "Stabilization Properties of the Payroll Tax," University of Illinois, 1968. I would like to thank those responsible for my NDEA and Brookings Fellowships under which the dissertation was written. I would like to thank Professors A. J. Heins, T. Yancey, G. Judge, and F. Shupp at the University of Illinois, and Sidney Carroll, John Brittain, Michael D. McCarthy at Brookings for many helpful comments on the content of this paper. Also, thanks are due to the Brookings Econometric Model Project, which gave support for several drafts of this paper.

[1]For a thorough review and extensive bibliography on all the major approaches, see Stanley Lebergott, "The Shape of the Income Distribution," *American Economic Review*, Vol. 44, No. 3 (June 1959), pp. 328–47.

[2]For an analysis of the relation of size of incomes to other variables, see F. Gerald Adams, "The Size of Individual Incomes," *Review of Economics and Statistics*, Vol. 40, N. 4 (November 1958), pp. 390–98. For the determinants of income inequality, see Almod Al-Samarrie and Herman P. Miller, "State Differentials in Income Concentration," *American Economic Review* Vol. 57, No. 1 (March 1967), pp. 59–72, and D. J. Aigner and A. J. Heins, "On the Determinants of Income Equality," *ibid.*, pp. 175–84.

[3]By wages we mean wages and salaries. See the Appendix for definitions.

consequent income distributions."[4] Leibenberg[5] shows how to derive a distribution from a preceeding one if all incomes change in the same proportion, which may not be the case in practice, as Bowman noticed when she observed that there were ". . . differential effects of changes in the level of business activity on income receivers in different parts of the total income distribution."[6]

It is the basic hypothesis of this paper that such differential effects do exist, at least for the wage distribution. The purpose here is to measure the change in the wage distribution in a parametric way as a vector and to relate the change to the business cycle. Emphasis will be on problems of definition and estimation and on some implications of the differential effects. After estimating a particular form of the wage distribution, one very suitable and easy to manipulate analytically, the rate of change in the wage distribution is measured. Then the analytic relationship between the size distribution framework and the Lorenz curve and Gini coefficient constructs paves the way for a discussion of inequality implications.

Offered as determinants of the rate of change vector are the level of and changes in the unemployment rate and changes in Gross National Product and in the Consumer Price Index. After predicting the 1966 wage distribution parameters, a simulation indicates what happens to the wage distribution when all determinants fluctuate together. Finally, a discussion of some uses and possible extensions of the analysis of the wage distribution concludes the study.

1. Definition of the Rate of Change in the Wage Distribution

Define f_t as the relative frequency function of wage earners, or density function, for year t. More precisely, $f_t(w)$ is the number of wage earners earning the wage w that year, divided by the total number of wage earners. Further, define F_t as the cumulative relative frequency function of wage earners for year t, called the size distribution of wages, or merely the wage distribution. More precisely,

$$F_t(w) = \int_0^w f_t(y)\, dy. \tag{1}$$

The movement of F over time may be defined according to wage levels. In particular, for any wage level $\bar{w}$ for year t, call it $\bar{w}_t$, where the "bar" signifies a given level of a variable (and not its "mean") and t dates it, there corresponds a certain wage level for the year $t + 1$ such that

$$\bar{w}_{t+1} = e^r \bar{w}_t \tag{2}$$

and

$$F_{t+1}(\bar{w}_{t+1}) = F_t(\bar{w}_t), \tag{3}$$

where e is the base of the natural logarithms and r is a variable. Given two curves F_t and F_{t+1} there is associated with each curve wage levels $\bar{w}_t$ and $\bar{w}_{t+1}$ as in

[4]D. G. Champernowne, "A model of Income Distribution," *Economic Journal*, Vol. 63, No. 250 (June 1953), p. 349.

[5]Maurice Liebenberg, "Nomographic Interpolation of Income Size Distributions," *Review of Economics and Statistics*, Vol. 38, No. 3 (August 1956), pp. 258–72.

[6]Mary Jean Bowman, "A Graphical Analysis of Personal Income Distribution in the United States," *American Economic Review*, Vol. 35, No. 4 (September 1945), pp. 607–28, reprinted in American Economic Association, *Readings in the Theory of Income Distribution*, (Philadelphia: The Blakiston Company, 1951), pp. 72–99. The quote is from p. 80.

equation (3) at a given F-value, say $\bar{F}$. Then from equation (2) r is measured at $\bar{F}$ as $\Delta \ln \bar{w}_{t+1}$, where Δ is the first-difference operator and ln the natural logarithm function. By varying $\bar{F}$ from zero to one, r is defined and measured continuously over the entire wage distribution. For purposes of estimation, the variable r is compiled as a vector across time, with the number of observations on the derived vector depending upon the number of F-levels at which it is measured.

It is not possible to measure r with f instead of F because f may be flat over some ranges of w. Also, the largest value of f_t may be greater than the largest value of f_{t+1}, thus no horizontal measurement is available between the two curves. Assuming F continuous and $f > 0$ for all $w > 0$, as we shall, a horizontal measurement is always available with F.

2. Measuring the Rate of Change in the Wage Distribution Vector

To measure r, it is preferable for several reasons to relate F to w by an explicit function. For the purpose of measuring r, that function should comply with the following criteria: (a) the function should fit the observations well, (b) be of sufficient simplicity to be useful analytically, and describe three basic properties of the cumulative relative frequency of wages: (c) F must pass through the origin, (d) F must be monotonically increasing, and (e) F must be asymptotic to unity as wages increase.[7]

Figure 1 shows F plotted against wage levels, w (in thousands of current dollars), for the years 1947, 1951, 1957, and 1966. It is seen that F is approximately linear over its lower ranges up to an F-level of about 0.825, whereupon it tapers off to unity. It is preferable to use a given F-value to terminate the linear portion rather than a given w-value, because, for example, a w-value of 3.0 satisfies the 1947 but not the 1966 distribution (too little of the linear portion is included below $w = 3.0$ in 1966), and $w = 7.0$ satisfies the 1966 but not the 1947 distribution (too much of the non-linear portion of the 1947 distribution is below $w = 7.0$). For only four of the years 1947–1966 are data on F available for $w < 0.5$ (see Appendix for the data); thus the data limitations suggest that if a linear approximation is used for at least part of the wage distribution, it should begin at $w = 0.5$ and end at $F = 0.825$. For $F > 0.825$ a non-linear analytic form is needed that is asymptotic to unity. For $w < 0.5$ a straight line from the linear portion at $w = 0.5$ through the origin appears appropriate from the little data that is available.

Therefore, the following analytic form for F is proposed:

$$F_t = \begin{cases} c_t w & \text{if } 0 \leq w \leq 0.5 \qquad (4.1) \\ a_t + b_t w & \text{if } w \geq 0.5 \text{ and } F_t \leq 0.825 \qquad (4.2) \\ 1 - g_t w^{-D} & \text{if } F_t \geq 0.825 \qquad (4.3) \end{cases}$$

[7]Aitchinson and Brown give (a) and (b) as two of four criteria for statistical description of personal income distributions, and they would no doubt agree with (d) and (e). Criterion (c) does not hold for incomes other than wages because of the possibility of losses with other types of incomes. Their other two criteria deal with the generation theory behind the analytic form and the economic sense that can be attached to its parameters. We have circumvented both criteria by taking the distribution as given and placing the emphasis on the change in the distribution, without regarding the analytic form of the distribution as an end in itself. See J. Aitchinson and J. A. C. Brown, *The Lognormal Distribution*, University of Cambridge Department of Applied Economics, Monograph No. 5 (Cambridge: Cambridge University Press, 1957), p. 108.

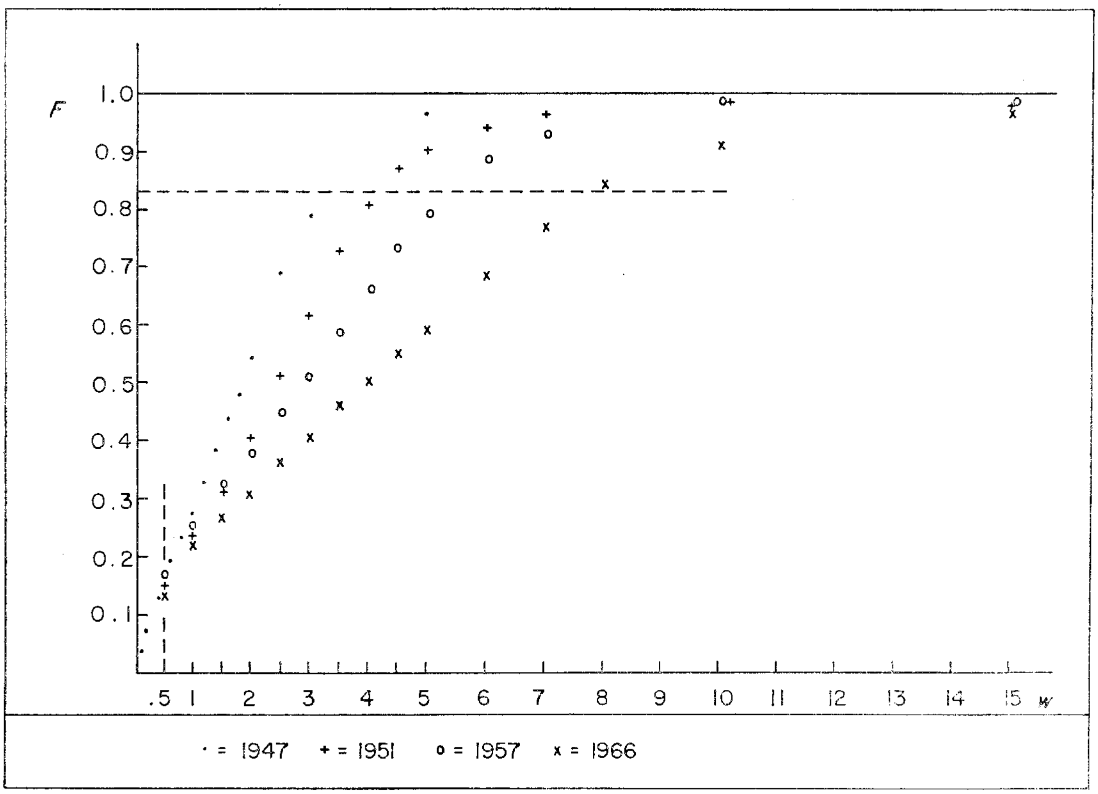

Figure 1. F for selected years
(Source: See Appendix)

where t is time and the coefficients c, a, b, and g are allowed to vary over time to accomodate the movement of the wage distribution. D is constant over time. Equation (4) is defined to be continuous, so that given a_t and b_t of the linear part (4.2), $c_t = (a_t + 0.5b_t)/0.5$ and $g_t = 0.175[(0.825 - a_t)/b_t]^D$. Therefore, given D, the wage distribution is completely defined if a_t and b_t of (4.2) are obtained.

To obtain a_t and b_t a simple linear regression appears appropriate. F should be regressed on w and not vice versa, because w is a non-stochastic variable while F is not, hence errors of observation associated with F will be accounted for in the regression $F_t = a_t + b_t w$. There is reason to suspect autocorrelation of the residuals in this equation (for each year), because if errors of observation on f are randomly distributed over wages, then F will have autocorrelated residuals because it sums f over wages, accumulating the random error terms.

Table 1 presents the results of estimating the coefficients a and b and the first-order autocorrelation coefficient ρ (estimated to the nearest tenth) by the Dhrymes procedure,[8] for the period 1947–1965. (See the Appendix for the data.) The coefficient a' is the coefficient of the variable consisting of one's, so $a = a'/(1-\hat{\rho})$,

[8]Phoebus J. Dhrymes, "On the Treatment of Certain Recurrent Non-Linearities in Regression Analysis," *Southern Economic Journal*, Vol. 33, No. 2 (October 1966), pp. 187–96. For a summary of Dhrymes' technique and its extension to the nth order autoregressive case, see Henry J. Cassidy, "Maximum Likelihood Estimation in an nth Order Autoregressive Disturbance Model," *Southern Economic Journal*, Vol. 35, No. 3 (January 1969), pp. 263–64.

TABLE 1

STATISTICS FROM LINEAR REGRESSIONS OF $F = a + bw$, $0.5 < w$ and $F < 0.825$

Year	ρ	a'	b	t-statistic for a'	t-statistic for b	$\bar{R}^2$	$D - W$	a
1947	0.0	0.0266	0.2567	3.305	55.635	0.997	2.090	0.0266
1948	0.2	0.0313	0.2331	4.572	46.804	0.995	1.345	0.0391
1949	0.1	0.0484	0.2252	8.744	61.601	0.997	1.056	0.0538
1950	1.0		0.2183		19.536	0.933	1.712	0.0458
1951	0.0	0.0392	0.1915	4.387	54.109	0.998	1.269	0.0392
1952	0.0	0.0416	0.1856	4.107	46.266	0.997	1.012	0.0416
1953	0.0	0.0572	0.1708	8.277	69.474	0.998	1.510	0.0572
1954	0.1	0.0662	0.1657	11.277	72.356	0.998	1.525	0.0736
1955	0.1	0.0860	0.1538	18.984	87.496	0.999	1.775	0.0955
1956	0.1	0.0969	0.1422	18.469	76.863	0.998	2.138	0.1077
1957	0.1	0.0978	0.1381	24.656	98.777	0.999	2.630	0.1087
1958	0.1	0.1153	0.1303	31.044	99.567	0.999	2.628	0.1281
1959	0.2	0.1030	0.1228	28.426	87.143	0.999	2.410	0.1287
1960	0.2	0.1080	0.1183	32.347	91.069	0.999	2.585	0.1351
1961	0.2	0.1157	0.1108	42.120	116.167	0.999	2.307	0.1446
1962	0.3	0.1044	0.1067	41.429	109.374	0.999	3.276	0.1491
1963	0.3	0.1035	0.1034	36.463	94.058	0.999	2.565	0.1479
1964	0.4	0.0941	0.0965	23.458	61.625	0.997	1.540	0.1568
1965	0.4	0.0907	0.0937	24.683	65.364	0.997	2.196	0.1511

except for 1950 when $\hat{\rho} = 1$ and there is no constant term. Here a equals the average of F less b times the average of w. The Durbin-Watson has no significance for us since we have accounted for autocorrelation, but it is reported for completeness.

It is well known that the Pareto density function fits the upper income brackets quite well, and its form (for wages) is $f(w) = Aw^{-(V-1)}$, where A and V are parameters and $V > 2$. Integrating $f(w)$ we have $F(w) =$ a constant $- [A/(V-2)]\, w^{-(V-2)}$. Now the constant equals one since F approaches one as w approaches infinity. Letting $D = V - 2$, we write $F(w)$ as

$$F = 1 - Aw^{-D}/D. \tag{5}$$

Let $w_0 = (0.825 - a)/b$ be the point on the w-scale of the intersection of (4.2) and (4.3). Substituting 0.825 and w_0 for F and w, respectively, in (5) we have $0.175 = Aw_0^{-D}/D$, and solving for A yields $A = 0.175\, Dw_0^{D}$. Substituting this for A in (5) yields

$$F = 1 - 0.175\,(w_0/w)^D. \tag{6}$$

Turning back to Table 1, ρ was estimated as less than 0.5 for all years except 1950. Although it is a subjective judgment that an autoregressive parameter of 0.4 is insignificant, we shall make that assumption. In 1950 only one observation is in the range $F > 0.825$, so we will assume away the autoregressive nature of the residuals for $F > 0.825$. In that case rearranging and taking logarithms of (6) yields

$$\ln\left(\frac{1-F}{0.175}\right) = D \ln(w_0/w), \tag{7}$$

which we shall use to estimate D for all the years, under the assumption that D is constant. There are 95 observations and we estimate (7) as

$$\ln\left(\frac{1-F}{0.175}\right) = \underset{(74.4)}{3.25}\ln(w_0/w) \tag{8}$$

$$\bar{R}^2 = 0.954 \qquad D - W = 2.135$$

We also present the relative frequency function implied by (4), which is derived by differentiating F with respect to w.

$$f_t = \begin{cases} c_t & \text{if } 0 \le w < 0.5 \quad (9.1) \\ b_t & \text{if } w \ge 0.5 \text{ and } F_t \le 0.825 \quad (9.2) \\ 3.25\, g_t w^{-4.25} & \text{if } F_t > 0.825 \quad (9.3) \end{cases}$$

As a matter of secondary interest, Figure 2 portrays the actual and estimated f for 1965. Since w is measured in thousands of dollars, the values of the actual f were adjusted to be comparable; i.e., if the length of a w-interval defining

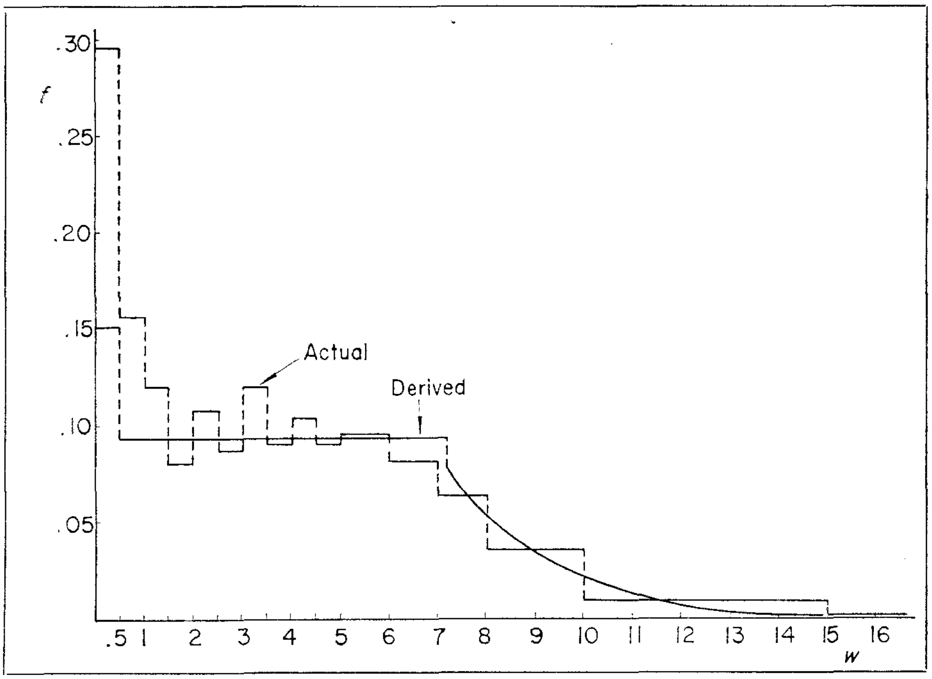

Figure 2. Actual and Derived f for 1965
(Source of Actual Data: See Appendix)

a value of f were 0.5, then that value of f is multiplied by 2. Except for the first two wage brackets ($w \le 1.0$) the estimated f appears to conform to the actual f. There appears to be a tendency for wages in a thousand-dollar bracket to be clustered above the integer values of 2, 3, and 4 rather than below them. It is not certain whether this phenomenon is due to a habit of employers to adjust wages upward to come out that way, or the tendency of respondents in the sample to inflate their wages above these integers when their wages are in fact just below them. However, it would complicate the analytic form of f and F to attempt to describe this phenomenon.

As an aside, the cofficient a has increased over the years 1947–1965, as shown in Table 1, which may be caused by the secular increase in part-time employment by dependent members of a family, who are included in the distribution statistics. It may also be caused by the part-time employment of the increasing number of retired people covered by social security. But this is a matter for further investigation.

We have formulated the movement of wage levels as a relative one. The F-value associated with a given value of w, say $\bar{w}$, is very different in 1947 than it was in 1966, as seen in Figure 1, which leads us to doubt that the r calculated at w would be significantly related to economic variables, because a completely different class of people, relatively speaking, earned $\bar{w}$ in 1947 than in 1966. We measure r, then, at given F-values and not at given w-values. Since we wish to compare F-curves for adjacent years we shall use the subscript notation referring to years t and $t + 1$. We now show that the analytic form of (4) allows r to vary over the linear range (4.2), but not over (4.1) or (4.3): r remains a constant over these ranges.

Let $F_t = F_{t+1} = \bar{F}$, say, where $\bar{F}$ is a given F-level, and measure r between years t and $t + 1$ at that F-level. In particular, $r = \ln(w_{t+1}/w_t)$. For (4.1) and year t, $w_t = \bar{F}/c_t$ and for year $t + 1$, $w_{t+1} = \bar{F}c_{t+1}$. Taking the ratio of w_{t+1} to w_t gives $w_{t+1}/w_t = c_t/c_{t+1}$, and $r = \ln(c_t/c_{t+1})$ in this range, which is free of $\bar{F}$, so that r remains constant over F under (4.1). For (4.2) $w_t = (\bar{F} - a_t)/b_t$ and $w_{t+1} = (\bar{F} - a_{t+1})/b_{t+1}$. Then $r = \ln[(\bar{F} - a_{t+1})b_t/\{(\bar{F} - a_t)b_{t+1}\}]$ which is not free of F and thus varies over values of F in this range. For (4.3)$w_t = [g_t/(1 - \bar{F})]^{1/3.25}$ and $w_{t+1} = [g_{t+1}/(1 - \bar{F})]^{1/3.25}$. Then $r = \ln[(g_{t+1}/g_t)^{1/3.25}]$, which is free of $\bar{F}$. Thus r is held at its value for $F = 0.825$ for the range $F \geq 0.825$.

We measured r at intervals of F of 0.05 from $F = 0.20$ through $F = 0.80$, which, for our data, complies with $w \geq 0.5$ and $F \leq 0.825$ for 1947 through 1965; thus r becomes an 18×1 vector (across time) with 13 observations corresponding to the 13 F-levels chosen. The values of r appear in the Appendix in Table 8.

3. The Relation of F to the Lorenz Curve and Wage Inequality

How does the size distribution relate to measures of inequality? In particular, how does it relate to the Lorenz Curve and the Gini coefficient? In equation (4) w was related to F as $F = P(w)$ for the wage distribution function. For the Lorenz curve F is related to a function of w, call it $A = Q(w)$ where

$$Q(w) = \int_0^w yf(y)\,dy \Big/ \int_0^\infty yf(y)\,dy$$

the proportion of aggregate wages under a given wage w.[9] Q is a strictly increasing function of w since we assume $f > 0$ for all $w > 0$, so we may write

[9]Aggregate wages equal $\int_0^\infty yLf(y)\,dy = L\int_0^\infty yf(y)\,dy$, where L is the number of people earning wages that year. L cancels out in the ratio $Q(w)$.

$w = Q^{-1}(A)$. The Lorenz curve may be written, then, as $F = P[Q^{-1}(A)]$, where F goes from zero to one as A does.

The Gini coefficient, G, is the shaded area in Figure 3 between the curve and the "line of equality":

$$G = \int_{A=0}^{1} P[Q^{-1}(A)]\, dA - 0.5. \tag{10}$$

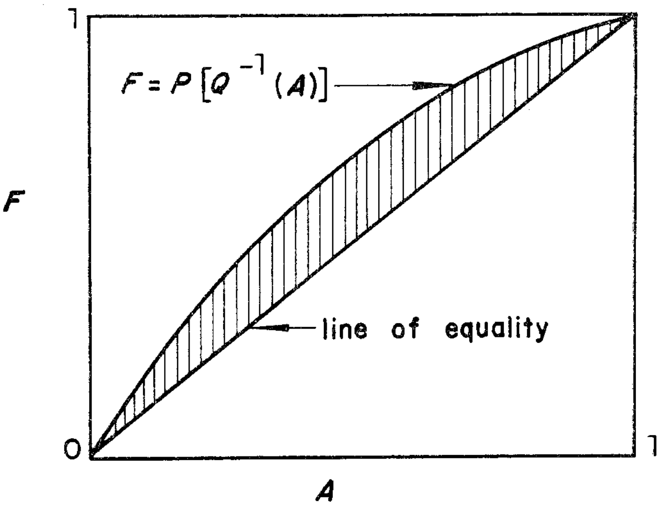

Figure 3. Lorenz curve (with axes reversed)

As G increases, inequality increases, according to Gini. Table 2 presents the analytic forms of $P(Q^{-1})$ for the three ranges of F, where $w_0 = (0.825 - a)/b$, the point where the linear and Pareto portions of F intersect, and

$$z = \int_0^{\infty} yf(y)\, dy,$$

the average wage.

From the table we calculate G as[10]

$$\begin{aligned} \mathrm{G} = {} & -0.5 + [(0.25)^{3/2}(c^2 - b^2)/3 - 0.125(ab + c - b) + z \\ & - (1 - a)bw_0^2/2 + b^2 w_0^3/3 - 3.25 g^{4.5/2.25}/5.5\, w_0^{5.5}]/z. \end{aligned} \tag{11}$$

[10]Let h be the proportion of the labor force who do not work for the entire year, which we indicate in the next section to be quite small, and let $\hat{F}(w)$ be the wage distribution of all those earning wages in addition to those of the labor force who do not. Then

$$\hat{F}(w) = F(w) + h(1 - F(w)) = (1 - h)F(w) + h.$$

Then the Gini ratio for the new definition of the wage distribution is

$$\hat{G} = \int_0^1 P[Q^{-1}(A)]\, dA - 0.5 = (1 - h)G + h - 0.5.$$

If h remains relatively constant, then $\hat{G}$ and G are in one-to-one relationship over time. Hence, using the concept of h changes the level of inequality, but not the change in inequality over time. Since the change in inequality has more intuitive meaning than the level, it is not necessary to use $\hat{F}$ instead of F.

TABLE 2
THE LORENZ CURVE $F = P[Q^{-1}(A)]$

Range of F	$F = P[Q^{-1}(A)]$
(i) $w < 0.5$	$(2czA)^{1/2}$
(ii) $w \geq 0.5$ and $F \leq 0.825$	$a + [2b\{zA - 0.125\,(c - b)\}]^{1/2}$
(iii) $F > 0.825$	$1 - g[(2.25/3.25g)\{-zA + (3.25/2.25)\,gw_0^{-2.25} + (b/2)w_0^2 + 0.125\,(c - b)\}]^{3.25/2.25}$

Table 3 presents the values of G for the years 1947 through 1965. Morgan appears to be quite correct when he claims that "short run fluctuations of income are less important than one might think in their effects on overall inequality measures."[11] However, given a sizeable short-run change in the wage distribution of the type examined in this paper, the Lorenz curve appears to change much less because of the transformation which condenses the wage distribution into the Lorenz curve, and an overall inequality measure such as the Gini coefficient shows even less change due to its aggregative nature.

TABLE 3
GINI CONCENTRATION RATIO FOR WAGES, 1947–1965

1947	0.200	1957	0.222
1948	0.203	1958	0.228
1949	0.207	1959	0.228
1950	0.205	1960	0.230
1951	0.204	1961	0.233
1952	0.204	1962	0.234
1953	0.209	1963	0.234
1954	0.213	1964	0.237
1955	0.219	1965	0.236
1956	0.222		

4. THE DETERMINANTS OF THE RATE OF CHANGE IN THE WAGE DISTRIBUTION VECTOR

Figure 1 shows that the wage distribution has shifted considerably over the period 1947–1966. Since wages are measured in current dollars, much of this movement is due to inflation, which is represented by the rate of change in the Consumer Price Index (CPI). Another reason for the shift is that productivity and *per capita* GNP have increased, and these changes are reflected in the changes in real GNP, henceforth GNP. As discussed below, the unemployment rate, u, is also important in explaining year-to-year changes in the wage distribution. Each of these variables may affect the wage distribution differently for different F-levels. The following equation is the result of testing these conjectures:[12]

[11] J. Morgan, "The Anatomy of Income Distribution," *Review of Economics and Statistics*, Vol. 44, No. 3 (August 1962), pp. 270–83. The quote is from page 270.

[12] With suitable matrix notation we may explain how we estimated equation (12). Let

$$\underline{r}' = [r'_{0.20}\, r'_{0.25} \ldots r'_{0.80}],$$

where the underlining denotes the vector as we have defined it, the prime denotes transposition, and $r_{0.20}$ is the 18×1 observation vector on the r variable for an F-level of 0.20, $r_{0.25}$ for an F-level of 0.25, and so on. We shall explain the notation for u, and the same explanation applies to the other independent variables. Let

$$\underline{u}' = [u'\, u' \ldots u'],$$

where u is the 18×1 observation vector covering the years 1948 through 1965. Let

$$\underline{u/F}' = [u'/0.20\ \ u'/0.25 \ldots u'/0.80],$$

$$
(12)\quad
\begin{aligned}
r = & - \underset{(2.562)}{0.0279} + \underset{(4.417)}{1.0595}\, u - \underset{(7.418)}{0.4770}\, u/F + \underset{(2.857)}{0.6638}\, \Delta u \\
& - \underset{(2.274)}{0.2026}\, \Delta u/F + \underset{(11.384)}{0.9103}\, \Delta \ln(\mathrm{GNP}) + \underset{(2.183)}{0.660}\, \Delta \ln(\mathrm{CPI}) \\
& + \underset{(4.983)}{0.2199}\, \Delta \ln(\mathrm{CPI})/F + \underset{(6.799)}{0.1453}\, \Delta \ln(\mathrm{CPI}_{-1})/F,
\end{aligned}
$$

$$\bar{R}^2 = 0.727 \qquad D - W = 2.810$$

The "t"-statistics are in parentheses below the coefficients; the data are described in the Appendix. The coefficients of $\Delta \ln(\mathrm{GNP})/F$ and $\Delta \ln(\mathrm{CPI}_{-1})$ had "t"-statistics of less than one and were omitted from the equation. The variable $1/F$ does not appear by itself because the multicollinearity problem then made it impossible to distinguish the differential effect each of the other variables has over different F-values.

If unemployment is high we would expect the wage distribution (F) to shift to the left over the lower ranges, and to the right over the higher ranges. If a person is registered as unemployed in a given year, it means he was unemployed for only part of that year, but most likely not all of the year. We verify this statement by turning to Table 4, which gives the percentage distribution of unemployment by duration of unemployment. As may be seen, an extremely high percentage of unemployment is registered by those unemployed less than five weeks, and the highest percentage in the category "27 weeks and over" is 15.3; so the unemployment duration distribution appears to be skewed toward a mode of somewhat less than five weeks. What this means, then, is that if a person normally making \$6,000 per year is unemployed for two months, he registers in that year's wage distribution at \$5,000; there is one less person at \$6,000 that year than there would have been had he been employed, and one more at \$5,000. For many individuals, it implies that r is negative over lower ranges of F and positive over higher ranges if unemployment is high. Equation (12) portrays this phenomenon. For example, let $u = 0.05$ and evaluate the quantity $1.0595\, u - 0.4770\, u/F$. At $F = 0.20$ the quantity is -0.66, a shift to the left, and at $F = 0.80$ it is $+0.03$, a shift to the right.

For the change in the unemployment rate we would expect the same type of response over F as with the unemployment rate, but with a lesser absolute effect, because the main effect comes from the existence of the skewed unemployment duration distribution and changes in unemployment change the skewness but little. Hence, the coefficients on Δu and $\Delta u/F$ are smaller than on u and u/F, respectively, as well as the fact that the values of Δu are much smaller than those of u itself.

As GNP has increased in the postwar period, two phenomena have occurred: (a) per capita real GNP has increased 48 per cent between 1947 and 1965 and (b)

---

then equation (12) may be written as

$$
\begin{aligned}
\underline{r} = & \; a_0 \underline{1} + a_1 \underline{u} + a_2 \underline{u/F} + a_3\, \underline{\Delta u} + a_4\, \underline{\Delta u/F} + a_5\, \underline{\Delta \ln\,(\mathrm{GNP})} \\
& + a_6\, \underline{\Delta \ln(\mathrm{CPI})} + a_7\, \underline{\Delta \ln(\mathrm{CPI})/F} + a_8\, \underline{\Delta \ln(\mathrm{CPI}_{-1})/F}.
\end{aligned}
$$

TABLE 4

PERCENTAGE DISTRIBUTION OF THE UNEMPLOYED, BY DURATION OF UNEMPLOYMENT, 1947–1966
(Persons 16 years of age and over)

Year	Less than 5 weeks	27 weeks and over	Year	Less than 5 weeks	27 weeks and over
1947	52.4	7.1	1957	49.3	8.4
1948	57.2	5.1	1958	38.1	14.5
1949	48.3	7.0	1959	42.4	15.3
1950	44.1	10.9	1960	44.6	11.8
1951	57.2	6.7	1961	38.3	17.1
1952	60.2	4.5	1962	42.4	15.0
1953	62.2	4.3	1963	43.0	13.6
1954	45.5	9.0	1964	44.8	12.7
1955	46.8	11.8	1965	48.4	10.4
1956	51.3	8.4	1966	53.4	8.4

Source: U.S., Bureau of Labor Statistics, *Handbook of Labor Statistics*, 1967 (Washington: U.S. Government Printing Office, 1967), p. 89.

productivity has increased from an index of 69.1 in 1947 to 125.5 in 1965.[13] Thus, as GNP increases, the wage distribution shifts to the right, indicating that wage earners are receiving a share of the increased GNP, roughly 91 per cent of the increase.[14] The effect is relatively the same over all the wage distribution, relative because r is measured as the difference in logarithms, rather than in absolute differences. To have r increase the same over all F implies that the labor force mix, the percentage of people employed at given wage levels, is preserved when output increases. For a period as short as a year, this should be expected.

We expect r to increase over all ranges of F in response to increased prices to maintain real wages. CPI was used for prices because it relates more directly to purchases by wage-earners than does the money GNP deflator, for example. It is conceivable that there would be a lag in the response of wage-earners to demand cost-of-living raises in response to price increases, because the money illusion washes away over time as people feel that their increased wages were not raised enough in relation to prices. Why the increase in r should be larger over the lower ranges of F as opposed to the higher ranges is not clear unless, that is, the lower wage earners spend a higher proportion of their wages than do the higher wage earners. As an example of the mechanism involved in equation (12), let us suppose that the CPI changed at the rates of 0, 0.01, and 0.01 for 3 consecutive years. In the first year, r would not be affected by the CPI, but in the second year it would. In fact, the average wage for the second year increases by slightly more than 1 per cent over the first year's. Then in the third year the average wage increases about 1.4 per cent in response to 1 per cent changes in CPI and CPI_{-1}. But the CPI_{-1} was already accounted for in year two. If we use the constant

[13]The productivity series is taken from U.S. Bureau of Labor Statistics, *Handbook of Labor Statistics*, 1967 (Washington: U.S. Government Printing Office, 1967), p. 107, Table 71. The series refers to total private output per man-hour, establishment basis. (The figures are practically the same for the labor force data.) The sources for the *per capita* GNP calculation (actually, GNP per labor force population) in (a) are quoted in the Appendix.

[14]We shall use the convention that when we say "percentage change" we mean the change in logarithms.

term we may adjust the result downward, if desired, under the hypothesis that there is an exponential time trend associated with movements in CPI. However, one could argue that wages do increase relatively more than does the CPI as part of the wage-price spiral or because of increased productivity, but this is a matter for speculation or further investigation.[15]

5. Prediction and Simulation

To describe the movement of the wage distribution all that needs to be shown is how the linear portion $F = a + bw$ changes. Let the subscript t refer to the year and the subscripts 1 and 2 refer to two given F-levels, F_1 and F_2. For the two levels 1 and 2 we write

$$F_1 = a_{t+1} + b_{t+1}[(F_1 - a_t)/b_t]\, e^{r1} \tag{13}$$

and

$$F_2 = a_{t+1} + b_{t+1}\, [(F_2 - a_t)/b_t]\, e^{r2} \tag{14}$$

where r_1 and r_2 are the measures of the horizontal movements of the line between years t and $t + 1$ at F-levels F_1 and F_2. Solving (14) for a_{t+1}, substituting for a_{t+1} in (13), and rearranging yields

$$b_{t+1} = \frac{b_t(F_1 - F_2)}{(F_1 - a_t)e^{r1} - (F_2 - a_t)e^{r2}}\,, \tag{15}$$

The coefficient a comes from either (13) or (14), after substituting for b_{t+1} of equation (15). Given the actual series of u, GNP and CPI through 1966, b and a are computed as 0.0874 and 0.1527 for predictions for 1966. On the other hand, the actual values are 0.0909 and 0.1390.[16]

From an observation of one, it is not discernible whether the rate of change in the wage distribution may be used as a reliable predictor. The predicted value of b misses the actual by 3.74 per cent and the value of a is off by 9.86 per cent of the actual. With the actual values of b and a, $G = 0.234$, and with the predicted values, $G = 0.236$, so that while actual inequality decreased (compared to the value of G in 1965 of 0.2355), predicted inequality increased. The non-linearities involved and the above results with the Gini coefficient would indicate that if it is used as a predictor, caution should be taken in interpreting the results.

However, using sets of values of u, GNP, and CPI should be helpful in understanding what happens to the wage distribution and inequality when different economic conditions occur. When comparing extreme economic conditions such as boom or bust, the prediction error becomes less significant in accounting for the differences. Consequently, the five hypothetical economic conditions in Table 5 give several extreme and intermediate economic possibilities and will serve as the basis for a simulation.

[15]The coefficients of CPI and CPI_{-1} may be biased because of the simultaneity of the relationship. (The same remark applies to the coefficient of GNP.) In this paper we merely bow to the problem, assuming GNP and CPI to be exogenous.

[16]The "actual" coefficients come from Dhrymes procedure:

$$\begin{aligned} F = {} & \underset{(28.504)}{0.0834} + \underset{(79.612)}{0.0909}\, w \\ R^2 = {} & 0.998 \qquad \mathrm{D-W} = 2.729 \end{aligned}$$

and $a = a'/(1 - \hat{\rho}) = 0.1390$ and $\hat{\rho} = 0.4$.

TABLE 5
HYPOTHETICAL ECONOMIC CONDITIONS

Case	Label		Δ ln(GNP)	Δln(CPI)	$\%\Delta L$	u	Δu
(1)	Steady growth		0.052	0.00	0.0181	0.040	0
(2)	Boom	(a)	0.100	0.01	0.0526	> 0.040	< 0 if less than full employment
		(b)	0.052	0.08	0.0181	0.035	0 if fully employed
(3)	Recession		− 0.050	− 0.03	− 0.0175	≥ 0.040	> 0
(4)	Retardation		0.000	0.002	0.000	≥ 0.040	> 0
(5)	Steady growth with inflation		0.052	0.015	0.0181	0.036	0

The standard case, steady growth, was derived as follows. First, relate real GNP to L, the employed labor force, by least-squares regression, as

$$\begin{aligned} \%\Delta L &= 0.35099\, \Delta \ln(\text{GNP}) \\ &\quad (7.866) \\ \bar{R}^2 &= 0.623 \qquad D - W = 1.550, \end{aligned} \tag{16}$$

where $\%\Delta$ means percentage change. Equation (16) is estimated from annual data for the period 1947–1965. (See the Appendix.)

The second step is to find the percentage change in the potential labor force. From the *Handbook of Labor Statistics*, 1967[17] the average annual growth rate of the total labor force from 1962 through 1965 is 1.81 per cent. Since the most recent data are not comparable to prior data,[18] it appears best to use 1.81 per cent as an indication of the latest findings.[19] Given $\%\Delta L = 0.0181$, we solve (16) for q^*, that rate of growth in real GNP necessary to maintain full employment starting at full employment, as 0.052.

Case (2), boom conditions, is divided into two subcases, (a) and (b) as suggested by Keynesian theory. If the labor force is less than fully employed, real GNP can grow at a very high rate (here chosen at 0.100) with little inflation, because the increased GNP acts primarily to mop up unemployment. The value of $\%\Delta(L)$ comes from equation (16); unemployment will be decreasing under these assumptions. On the other hand, if the labor force is fully employed ($u = 0.04$), real GNP can grow no faster than the potential labor force allows it to grow; so inflation is a continuation of the boom conditions: money GNP grows but real GNP growth is limited by the growth in the labor force.

For Case (3), recession, a value of Δ ln(GNP) of about ten percentage points below that of Case (1) was chosen. From equation (16) we obtain $\%\Delta(L)$, and Δ ln(CPI) was, again, chosen arbitrarily, here to represent deflation. The value of Δ ln(CPI) of -0.03 is a relatively low one, being lower than any annual value since World War II. Unemployment increases, the rate of which is determined by $\%\Delta(L)$ and the initial conditions. Case (4) represents a situation not unlike the retardations of the post war period. We have assumed slight inflation, and

[17] *Ibid.*, p. 21. [18] *Ibid.*, p. 22, n. 1

[19] It is interesting to note that in the U.S., Department of Labor, *Manpower Report of the President, 1967* (Washington: U.S. Government Printing Office, 1967), p. 273, Table E-7, the estimate for the decade 1970–1980 is 18.1 per cent, having been derived, apparently, the same way as we derived the annual change.

as a consequence of $\%\Delta(L) = 0$, unemployment increases. Case (5), which combines growth with inflation, is very much like the growth of the actual paths GNP, CPI, and L for the period 1962–1965. The unemployment rate, unlike that period, however, is forced to 0.036, which presumably would occur if Case (5) started from a full employment situation.

Table 6 presents the values of a, b, and G for the fifth, sixth, and seventh years of simulation, under the following plan. Each case was run for five years and then each of the five cases ran for two more years for each of the five cases. The initial year is 1965, with $a = 0.1512$ and $b = 0.0938$, assuming full employment ($u = 0.04$) and no inflation in 1965.

In the simulation b decreases from its value in 1965 for all economic cases, and a increases in all cases except for Case (2), boom conditions. The Gini coefficient shows that for the boom, inequality decreases slightly. Otherwise,

TABLE 6

SELECTED VALUES OF a, b, AND G FOR THE SIMULATION

Case	Year	a	b	G
		Case (1) through year 5		
	5	0.1671	0.0754	0.241
(1)	6	0.1698	0.0723	0.242
(1)	7	0.1723	0.0693	0.243
(2)	6	0.1667	0.0700	0.241
(2)	7	0.1648	0.0640	0.241
(3)	6	0.1737	0.0776	0.243
(3)	7	0.1826	0.0787	0.245
(4)	6	0.1713	0.0743	0.242
(4)	7	0.1764	0.0724	0.244
(5)	6	0.1689	0.0721	0.242
(5)	7	0.1705	0.0686	0.242
		Case (2) through year 5		
	5	0.1416	0.0609	0.235
(1)	6	0.1429	0.0575	0.235
(1)	7	0.1454	0.0551	0.236
(2)	6	0.1396	0.0557	0.234
(2)	7	0.1378	0.0509	0.234
(3)	6	0.1469	0.0617	0.236
(3)	7	0.1554	0.0626	0.238
(4)	6	0.1445	0.0591	0.236
(4)	7	0.1494	0.0577	0.237
(5)	6	0.1420	0.0573	0.235
(5)	7	0.1437	0.0545	0.236
		Case (3) through year 5		
	5	0.2118	0.0903	0.251
(1)	6	0.2233	0.0791	0.255
(1)	7	0.2326	0.0692	0.259
(2)	6	0.2211	0.0776	0.255
(2)	7	0.2272	0.0676	0.258
(3)	6	0.2257	0.0856	0.255
(3)	7	0.2393	0.0800	0.259
(4)	6	0.2242	0.0816	0.255
(4)	7	0.2351	0.0730	0.259
(5)	6	0.2096	0.0936	0.250
(5)	7	0.2112	0.0890	0.251

TABLE 6

SELECTED VALUES OF a, b, AND G FOR THE SIMULATION (*cont.*)

Case	Year	a	b	G
		Case (4) *through year* 5		
	5	0.1871	0.0781	0.246
(1)	6	0.1943	0.0712	0.249
(1)	7	0.2008	0.0650	0.250
(2)	6	0.1917	0.0701	0.248
(2)	7	0.1940	0.0640	0.249
(3)	6	0.1973	0.0768	0.249
(3)	7	0.2085	0.0746	0.252
(4)	6	0.0955	0.0734	0.249
(4)	7	0.2038	0.0683	0.251
(5)	6	0.1864	0.0777	0.246
(5)	7	0.1880	0.0739	0.243
		Case (5) *through year* 5		
	5	0.1614	0.0732	0.239
(1)	6	0.1639	0.0698	0.240
(1)	7	0.1665	0.0669	0.241
(2)	6	0.1608	0.0676	0.240
(2)	7	0.1589	0.0618	0.240
(3)	6	0.1679	0.0749	0.241
(3)	7	0.1768	0.0760	0.243
(4)	6	0.1655	0.0718	0.241
(4)	7	0.1706	0.0700	0.242
(5)	6	0.1631	0.0696	0.240
(5)	7	0.1647	0.0622	0.241

wage inequality increases, most markedly in a depression (Case (3)), due to the high unemployment rate, which implies that although many workers may continue to make the same real wage (after accounting for deflation), a very large number are employed only part of the year, causing the standard deviation of wages (hence inequality) to increase.

6. USES AND EXTENSIONS

Both the constructs of the explicit form of the wage distribution and changes in it are of practical and theoretical interest. In addition to being able to indicate the level and change in wage inequality, they are useful for tax purposes. The payroll tax used to finance the Old-Age, Survivors, Disability, and Health Insurance programs has a structure consisting of a single (combined employer-employee) tax rate, p, applied to all wages below a wage ceiling wc.[20] For generality, suppose there is also an exemption level ex, below which no tax is paid. Then the payroll tax revenue T is defined by

$$T = p \,.\, L\left[\int_{ex}^{wc} yf(y)\,dy + wc\{1 - F(wc)\} - ex\{1 - F(ex)\}\right], \tag{17}$$

[20]Disregard the self-employed for this discussion.

where L is the average labor force for the year.[21] A useful exercise is to let $ex = 0$ and formulate an equal yield trade-off between wc and p, and then let wc approach infinity and find the equal yield trade-off between ex and p.

Figure 4 portrays these trade-offs for 1965, where T is defined by the 1965 tax structure of $wc = 4.8$ (thousands of dollars), $p = 0.0725$, and $ex = 0$. Turning to the $wc - p$ trade-off, a low wage ceiling is very expensive in terms of the tax rate because very little of everyone's wages is taxable. As wc becomes

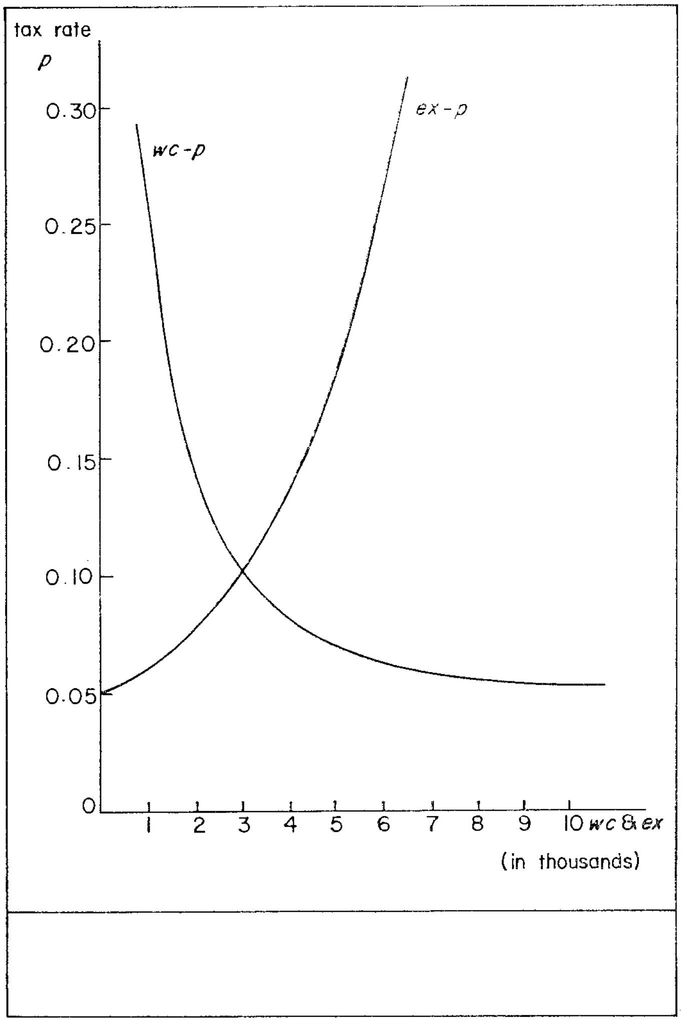

Figure 4. Equal yield $wc - p$ and $ex - p$ trade-offs for 1965

[21]It would not be very useful to use equation (17) and the rate of change vector to estimate payroll tax revenues, because equation (17) applies to a calendar year and an accrual basis, whereas the payroll tax revenue is reported on a fiscal year and cash basis. See revenue figures in U.S. Social Security Administration, *Social Security Bulletin*, Vol. 3, No. 11 (November 1967), p. 28, Table M-4. The Social Security Administration does not venture an accrual estimate, and the accrual basis series of the Department of Commerce is only an approximation, using Social Security data as the basis of their calculations. It is interesting to note, however, that the Current Population Survey covers very closely those people covered by social security today.

larger, around 6 to 7 thousand, any further increase in it yields little revenue (as indicated by the nearly horizontal slope of the $wc - p$ curve) above an F-level of around 0.75. This is not surprising since the only revenue gained comes from those people making more than wc, and then the additional revenue comes only from the tax rate applied to the wages just included by the increased wage ceiling. On the other hand, for the $ex - p$ trade-off, as ex increases from zero the tax rate must rise sharply, because not only are those earning ex relieved of paying tax when ex is increased, but also all those wage earners with wages above the new ex are exempt from paying tax on those wages excluded by the increase in ex.

Besides providing insights into the tax structure possibilities of a given year, the anlaysis may be extended to tax structure possibilities over time. In another study[22] the author has constructed equal yield lines showing the $ex - wc - p$ trade-offs, upon which was superimposed a map of equal tax elasticity (E_T) lines defined as $E_T = \Delta \ln(T)/\Delta \ln(W)$, where W is aggregate wages. The trade-offs were constructed for 1970, assuming steady growth (Case 1) from 1965. Then the tax elasticity map was constructed for 1970–1971 assuming steady growth through 1971. If a tax structure on the iso-elasticity line of $E_T = 1$ were chosen for 1970 and 1971, then E_T remains one (for 1970–1971) for all practical purposes if any of the other four hypothetical economic cases are assumed for 1971. However, if a tax structure is chosen that corresponds to an E_T different from one for steady growth, then E_T does not remain the same for the other economic circumstances, and in several cases E_T becomes "perverse" in terms of stabilizing the economy.[23]

The extensions of the analysis of the rate of change in the wage distribution vector are numerous. Other incomes should be tested, in which case different analytical forms for the distributions probably need to be found, and different determinants of the rate of change vector investigated, but the method of approach is clear. Disaggregation within each income base may be useful, such as deriving separate male and female income distributions. No doubt the female wage distribution responds more readily than the male one to economic changes, but that is an hypothesis to be tested. The earning unit used in the wage distribution statistics was the individual wage earner, but for welfare and tax purposes, for example, having the family as the basic earning unit may be more informative. Finally, a combination of income distributions could be incorporated into a simultaneous equation framework. It is hoped that research in these areas will be forthcoming in the near future, so that the state of size distribution theory may be advanced.

[22]See the author's unpublished Ph.D. dissertation "Stabilization Properties of the Payroll Tax," University of Illinois, 1968.

[23]A copy of the analysis is available upon request.

APPENDIX

The Data

Table 7 gives the observed values of F, the cumulative relative frequency of wages and salaries of persons 14 years old and over. The data are taken from various issues of U.S. Bureau of the Census, *Current Population Reports*, Series

P-60.[1] The subscripts on w refer to the years over which that w applies. The round-off error was adjusted so that $F = 1.0$ when w approaches infinity, for each of the years 1947 through 1966.

The population covered by the current population survey excludes inmates of institutions and includes only those members of the Armed Forces living off post or with families on post.[2] The definition used for wages and salaries is those "earnings received for work performed as an employee during the calendar year. . . . It includes wages, salary, Armed Forces pay, commissions, tips, piece-rate payments, and cash bonuses earned, before deductions were made for taxes, bonds pensions, union dues, etc."[3] Thus we see that the current population survey covers very closely those people and wages subject to the payroll tax.

TABLE 7

OBSERVED VALUES OF F

(Wages, w, in thousands of dollars)

$w_{47,50}$	F_{47}	F_{50}	w_{48}	F_{48}	w_{49}	F_{49}	$w_{51,52}$	F_{51}
5.0	0.966	0.944	10.0	0.996	10.0	0.996	15.0	0.990
3.0	0.787	0.699	6.0	0.981	7.0	0.990	10.0	0.983
2.5	0.690	0.595	5.0	0.958	6.0	0.980	7.0	0.965
2.0	0.541	0.464	4.5	0.939	5.0	0.959	6.0	0.945
1.8	0.480	0.419	4.0	0.904	4.5	0.937	5.0	0.899
1.6	0.440	0.388	3.5	0.845	4.0	0.899	4.5	0.866
1.4	0.383	0.347	3.0	0.745	3.5	0.838	4.0	0.805
1.2	0.326	0.309	2.5	0.640	3.0	0.738	3.5	0.723
1.0	0.277	0.274	2.0	0.495	2.5	0.628	3.0	0.614
0.8	0.234	0.239	1.8	0.445	2.0	0.496	2.5	0.516
0.6	0.191	0.200	1.6	0.405	1.8	0.447	2.0	0.405
0.4	0.127	0.146	1.5	0.377	1.6	0.412	1.5	0.315
0.2	0.069	0.087	1.4	0.361	1.5	0.381	1.0	0.238
0.1	0.036	0.050	1.2	0.314	1.4	0.365	0.5	0.146
			1.0	0.271	1.2	0.320		
			0.8	0.232	1.0	0.278		
			0.6	0.192	0.8	0.241		
			0.5	0.156	0.6	0.198		
			0.4	0.131	0.5	0.168		
			0.2	0.063	0.4	0.142		
			0.1	0.019	0.2	0.085		
					0.1	0.048		

[1]The specific sources in the P-60 series of the *Current Population Reports* from which the data are derived are listed below. We shall give the issue number and the date of issue, in parentheses, followed by the page number and table number, for the years 1947 through 1966, respectively; e.g., the first entry is written as 5 (February 7, 1949), 28, 21, which means that for 1947, the issue was No. 5, February 7, 1949 was the date of publication, and page 28, Table 21 is the specific source. The sources are 5 (February 7, 1949), 28, 21; 6 (February 14, 1950), 28, 18 and 29, 20 (the data for 1948 and 1949 comes from two tables, which we integrated into one series for each year); 7 (February 18, 1951), 35, 23 and 36, 25; 9 (March 25, 1952), 38, 23; 11 (May, 1953), 32, 12; 14 (December 31, 1953), 22, 11; 16 (May, 1955), 23, 10; 19 (October, 1955), 23, 10; 23 (November, 1956), 23, 10; 27 (April, 1958), 46, 27; 30 (December, 1958), 45, 27; 33 (January 15, 1960), 50, 39; 35 (January 5, 1961), 51, 36; 37 (January 17, 1962), 53, 36; 39 (February 28, 1963), 43, 40; 41 (October 21, 1963), 50, 28; 43 (September 29, 1964), 49, 32; 47 (September 24, 1965), 50, 32; 51 (January 12, 1967), 45, 32; and 53 (December 28, 1967), 50, 32.

[2]U.S., Bureau of the Census, *Current Population Reports*, Series P-60, No. 51, "Income in 1965 of Families and Persons in the United States" (Washington: U.S. Government Printing Office, 1967), p. 7.

[3]*Ibid.*, p. 9.

TABLE 7

OBSERVED VALUES OF F (*cont.*)

(Wages, w, in thousands of dollars)

F_{52}	w_{53-58}	F_{53}	F_{54}	F_{55}	F_{56}	F_{57}	F_{58}	w_{59-65}
0.998	25.0	0.999	0.999	0.999	0.999	0.999	0.999	25.0
0.993	15.0	0.997	0.997	0.997	0.996	0.996	0.995	15.0
0.975	10.0	0.990	0.989	0.989	0.985	0.983	0.979	10.0
0.954	7.0	0.966	0.964	0.956	0.944	0.938	0.926	8.0
0.896	6.0	0.937	0.934	0.921	0.901	0.888	0.871	7.0
0.853	5.0	0.869	0.866	0.844	0.813	0.794	0.776	6.0
0.791	4.5	0.822	0.816	0.791	0.760	0.738	0.723	5.0
0.705	4.0	0.749	0.746	0.715	0.679	0.663	0.650	4.5
0.591	3.5	0.667	0.663	0.637	0.608	0.595	0.586	4.0
0.499	3.0	0.563	0.562	0.547	0.526	0.516	0.512	3.5
0.395	2.5	0.479	0.480	0.477	0.460	0.456	0.455	3.0
0.310	2.0	0.384	0.397	0.396	0.384	0.380	0.384	2.5
0.234	1.5	0.312	0.322	0.330	0.358	0.320	0.328	2.0
0.149	1.0	0.238	0.247	0.256	0.325	0.251	0.262	1.5
	0.5	0.146	0.151	0.163	0.166	0.163	0.178	1.0
								0.5

F_{59}	F_{60}	F_{61}	F_{62}	F_{63}	F_{64}	F_{65}	F_{66}
0.998	0.998	0.997	0.998	0.997	0.997	0.997	0.996
0.993	0.992	0.989	0.990	0.989	0.987	0.985	0.980
0.974	0.969	0.961	0.958	0.952	0.940	0.936	0.917
0.944	0.934	0.922	0.914	0.901	0.883	0.865	0.839
0.907	0.893	0.875	0.865	0.845	0.824	0.802	0.772
0.843	0.827	0.805	0.788	0.767	0.745	0.721	0.688
0.743	0.724	0.702	0.683	0.664	0.648	0.625	0.595
0.689	0.674	0.650	0.634	0.619	0.601	0.580	0.555
0.619	0.609	0.589	0.577	0.564	0.547	0.528	0.504
0.560	0.552	0.535	0.525	0.515	0.497	0.483	0.461
0.489	0.485	0.472	0.464	0.452	0.437	0.423	0.403
0.434	0.430	0.423	0.417	0.404	0.391	0.380	0.363
0.368	0.366	0.363	0.357	0.348	0.339	0.326	0.312
0.316	0.316	0.314	0.310	0.303	0.295	0.286	0.270
0.251	0.253	0.251	0.243	0.241	0.232	0.226	0.212
0.168	0.167	0.170	0.163	0.161	0.151	0.148	0.137

The Social Security Administration has (unpublished) size distributions of wages for the same period, but they appear less useful than the current population survey data for two reasons. Since the social security data are derived from taxable wage reports only, estimates of wages above the wage ceiling are needed. Also, the rapid and continual expansion of social security since its inception means that the new population sectors brought into the system may make some years' data not comparable to other years'.

Another source of wage distributions which was not used is from the Internal Revenue Service, the main reason being that their wages are distributed according to adjusted gross income classes instead of wage classes. (This source would be appropriate if we were to study the revenue response of substituting the income tax for the payroll tax on wages.) The current population survey distributions, then, appear to be the most useful of available wage distributions.

Table 8 presents the calculated values of r, the rate of change in the wage distribution, F, for values of F from 0.20 through 0.80 at intervals of 0.05.

TABLE 8

CALCULATED VALUES OF r FOR SELECTED F-LEVELS

Year $t+1$	$F = 0.20$	$F = 0.25$	$F = 0.30$	$F = 0.35$	$F = 0.40$	$F = 0.45$	$F = 0.50$
1948	0.0214	0.0387	0.0495	0.0570	0.0624	0.0665	0.0697
1949	−0.0614	−0.0379	−0.0236	−0.0140	−0.0072	−0.0020	0.0020
1950	0.0847	0.0713	0.0633	0.0580	0.0541	0.0513	0.0490
1951	0.1729	0.1627	0.1565	0.1523	0.1493	0.1470	0.1453
1952	0.0166	0.0201	0.0223	0.0238	0.0249	0.0257	0.0264
1953	−0.0214	0.0046	0.0202	0.0307	0.0381	0.0437	0.0480
1954	−0.0911	−0.0581	−0.0392	−0.0269	−0.0183	−0.0120	−0.0071
1955	−0.1167	−0.0588	−0.0279	−0.0086	0.0045	0.0141	0.0213
1956	−0.0453	−0.0036	0.0171	0.0295	0.0377	0.0435	0.0479
1957	0.0187	0.0224	0.0241	0.0252	0.0259	0.0263	0.0267
1958	−0.1813	−0.0899	−0.0491	−0.0260	−0.0110	−0.0006	0.0071
1959	0.0504	0.0542	0.0557	0.0566	0.0571	0.0575	0.0578
1960	−0.0558	−0.0164	−0.0004	0.0082	0.0136	0.0173	0.0200
1961	−0.0935	−0.0213	0.0057	0.0199	0.0286	0.0345	0.0388
1962	−0.0478	−0.0061	0.0084	0.0157	0.0202	0.0231	0.0253
1963	0.0548	0.0432	0.0393	0.0373	0.0361	0.0353	0.0347
1964	−0.1182	−0.0219	0.0091	0.0243	0.0334	0.0395	0.0438
1965	0.1534	0.0885	0.0682	0.0582	0.0523	0.0484	0.0456

Year $t+1$	$F = 0.55$	$F = 0.60$	$F = 0.65$	$F = 0.70$	$F = 0.75$	$F = 0.80$
1948	0.0723	0.0744	0.0762	0.0777	0.0790	0.0802
1949	0.0052	0.0079	0.0101	0.0119	0.0135	0.0149
1950	0.0473	0.0458	0.0446	0.0436	0.0427	0.0419
1951	0.1438	0.1427	0.1417	0.1409	0.1401	0.1395
1952	0.0269	0.0273	0.0276	0.0279	0.0282	0.0284
1953	0.0515	0.0543	0.0567	0.0587	0.0604	0.0619
1954	−0.0032	−0.0000	0.0026	0.0048	0.0067	0.0083
1955	0.0270	0.0316	0.0353	0.0385	0.0412	0.0435
1956	0.0513	0.0540	0.0563	0.0581	0.0597	0.0610
1957	0.0270	0.0272	0.0274	0.0275	0.0277	0.0278
1958	0.0130	0.0176	0.0214	0.0246	0.0272	0.0295
1959	0.0580	0.0581	0.0583	0.0584	0.0585	0.0586
1960	0.0221	0.0237	0.0250	0.0261	0.0270	0.0277
1961	0.0420	0.0445	0.0466	0.0482	0.0496	0.0508
1962	0.0269	0.0281	0.0291	0.0299	0.0306	0.0312
1963	0.0343	0.0340	0.0337	0.0335	0.0333	0.0332
1964	0.0470	0.0495	0.0515	0.0531	0.0545	0.0556
1965	0.0435	0.0419	0.0406	0.0395	0.0387	0.0379

The data on employment and unemployment are taken from the *Handbook of Labor Statistics, 1967, op. cit.*, p. 21, Table 1. Real Gross National Product for 1947–1962 comes from the U.S. Office of Business Economics, *Survey of Current Business*, Vol. 45, No. 8 (August, 1965), p. 51, Table 16, and for 1963–1965 from the *Survey of Current Business*, Vol. 47, No. 7 (July, 1967), p. 13, Table 1.2. The consumer price index is that for all items, U.S. city average, taken from the *Handbook of Labor Statistics*, 1967, p. 200, Table 105.