AN ANALYSIS OF INCOME DISTRIBUTION IN THE NORTHEAST USING THE GAMMA DENSITY

by

Richard N. Boisvert

July 1977 No. 77-20

AN ANALYSIS OF INCOME DISTRIBUTION IN THE NORTHEAST USING THE GAMMA DENSITY*

Richard N. Boisvert

Despite the reduction during the 1960's in the number of families officially classified in poverty and the abandonment of many poverty programs, Schultz, among others, has argued that the public's demand for lessened income inequality has never been completely satisfied. The Administrations' responses to high unemployment, to recent rounds of inflation and to fiscal crises at all levels of government suggests that income distribution is of top priority. While those at the lower end of the income distribution should not be forgotten, persons in middle income categories are encountering problems not experienced in the rapid-growth, low-inflation years of the 1960's. Lifestyles built on increasing real incomes cannot be maintained when growth rates decline. Factors determining the distribution of the Nation's product must be identified if one hopes to understand the impacts of controlled economic expansion in the face of growing energy and environmental problems.

To date, most analysis of income distribution has focused on the relationship between aggregate economic performance and a summary measure of the distribution such as income inequality or the size of the low income population (e.g., Thurow, 1967; Boisvert). Notable exceptions by Salem and Mount, Metcalf and Thurow, 1970, relate state and national

---

* This paper was presented at the Annual Meetings of the American Agricultural Economics Association, San Diego, California, July 31 - August 3, 1977. Richard N. Boisvert is an associate professor of Agricultural Economics, Cornell University.

economic performance to the entire size distribution of income. Being able to describe the entire distribution in some convenient way enables one to examine a number of characteristics of the distribution simultaneously.

This paper extends these analyses in two ways. First, it is hypothesized that the size and functional distributions are jointly determined with aggregate economic performance. Second, a regional focus is motivated by the fact that lower future growth rates may increase the need for understanding how income distributions among regions can be altered.

Many economists argue that state and county units are not well-suited for the study of economic problems. This analysis utilizes 70 multi-county regions in the Northeast.[1/] Edwards and Coltrane believe that these regions approach functional economic areas and are aligned geographically with regional economic problems.

Describing the Income Distribution

A number of analytic distributions have been used to describe income distribution. Following Salem and Mount, the gamma density is used here and is defined for family income X (for $0 < X < \infty$; $\alpha > 0$; and $\lambda > 0$) as

$$(1)\quad f(X;\alpha,\lambda) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} X^{\alpha-1} e^{-\lambda X}; \text{ where } \Gamma(\alpha) = {}_0\!\int^{\infty} e^{-u} u^{\alpha-1}\, du.$$

Maximum likelihood estimates of the parameters can be obtained with the help of sample geometric mean ($\tilde{X}$) and arithmetic mean ($\overline{X}$): $\hat{\lambda} = \hat{\alpha}/\overline{X}$; and $\log \hat{\alpha} - \psi(\hat{\alpha}) = \log (\overline{X}/\tilde{X})$, where $\psi(\hat{\alpha}) = d \log \Gamma(\hat{\alpha})/d\hat{\alpha}$ (tabulated in Davis).

1/ States included are: Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Virginia and West Virginia. See Boisvert for delineation of the regions.

Since this is a two parameter distribution, the standard summary characteristics of the distribution can be written in terms of $\hat{\alpha}$ and $\hat{\lambda}$. For present purposes, the arithmetic mean ($\overline{X} = \hat{\alpha}/\hat{\lambda}$) and the coefficient of momental skewness ($\Sigma = 1/\sqrt{\hat{\alpha}}$) are of most interest. Mean income varies directly with $\hat{\alpha}$ and inversely with $\hat{\lambda}$. Skewness of the distribution depends only on $\hat{\alpha}$. That is,

(2) $\partial\overline{X}/\partial\hat{\alpha} = 1/\hat{\lambda}$; $\partial\overline{X}/\partial\hat{\lambda} = -\hat{\alpha}\hat{\lambda}^{-2}$; and $\partial\Sigma/\partial\hat{\alpha} = -1/2\hat{\alpha}^{-3/2}$.

To analyze income distribution in the Northeast, estimates of the parameters of gamma distributions (one for each of the 70 multi-county regions) were obtained for 1970 family income data (U.S. Department of Commerce). The performance (deviations of predicted from observed probabilities of families in 15 income categories) of the gamma and lognormal distributions were compared. The sums of squared deviations about the observed in each group were always higher (86% on average) using the lognormal.[2/]

Changes in important characteristics of the distribution as α and λ change depend on the initial values. Starting at average values of $\hat{\alpha}$ and $\hat{\lambda}$ across all regions, both average family income and skewness are responsive to changes in the parameters (table 1). Changes in average income are proportional changes in α. Decreases in λ result in a larger percentage increase in average income; increases in λ lead to a less than proportionate reduction in average income. The skewness measure varies

2/ The displaced lognormal and beta distributions may have fit the data a bit better (Metcalf; Thurow, 1970) but their parameters are difficult to interpret. The gamma is a good compromise.

Table 1. Income Distribution Response to Changes in Gamma Distribution Parameters

Item	$\hat{\lambda} - .1\hat{\lambda}$, $\hat{\alpha}$	$\hat{\lambda}$, $\hat{\alpha} - .1\hat{\alpha}$	$\hat{\lambda}$, $\hat{\alpha}$	$\hat{\lambda}$, $\hat{\alpha} + .1\hat{\alpha}$	$\hat{\lambda} + .1\hat{\lambda}$, $\hat{\alpha}$
Average income ($\overline{X}$)	\$11,465 (+11.0%)	\$9,298 (-10.0%)	\$10,328	\$11,364 (+10.0%)	\$9,451 (-8.5%)
Skewness (Σ)	0.693 (0)	0.730 (5.3%)	0.693	0.660 (-4.8%)	0.693 (0)
Families with incomes less than \$5,000, in % a/	22.8 (-12.6%)	27.2 (+4.2%)	26.1	21.7 (-16.9%)	29.3 (+12.3%)
Families with incomes between \$5,000-\$10,000, in % a/	28.6 (-7.7%)	30.0 (-3.2%)	31.0	30.0 (-3.2%)	33.0 (+6.5%)
Families with incomes greater than \$10,000, in % a/	48.6 (+13.3%)	42.8 (-0.2%)	42.9	48.3 (+12.6%)	37.7 (-12.1%)

Note: $\hat{\alpha} = 2.085$ and $\hat{\lambda} = 2.0187 \times 10^{-4}$, are the average estimated values for 1970 family income in 70 multi-county regions in the Northeast. The range in $\hat{\alpha}$ is from 1.591 to 2.447 and the range in $\hat{\lambda}$ is 1.0201×10^{-4} to 2.7291×10^{-4}. The numbers in parentheses are percent changes from the respective items' values for $\hat{\alpha}$ and $\hat{\lambda}$.

a/ These figures were calculated from a polynomial approximation to the standardized distribution function, redefining the cumulative density as

$$F(X;\alpha,\lambda) = \frac{1}{\Gamma(\alpha)} \int_0^X \lambda^\alpha Y^{\alpha-1} e^{-\lambda y}\, dy = I^*(u;\alpha-1);$$ where $u = X\lambda/\sqrt{\alpha}$ and I^* is the ratio of incomplete

to complete gamma functions (Pearson; Salem and Mount).

inversely with α. A 10% change in α leads to a less than proportionate change in the skewness of income distribution.[3/]

Perhaps more important than being able to estimate changes in these summary characteristics is the ability to identify changes in the proportion of families in various income categories. Relative importance of the income categories depends on the specific policy analysis. To illustrate, the impact on three income groups is examined. A 10% reduction in λ or a 10% increase in α (relative to $\hat{\alpha}$ and $\hat{\lambda}$) leads to approximately the same increase in the proportion of families with incomes greater than $10,000. One can also determine that the change in α would be more advantageous for families in the lowest income classes. The proportion of families with incomes less than $5,000 falls by 16.9% as α increases by 10% but by only 12.6% when λ decreases by 10%. A larger share of the increase in the proportion of families with incomes greater than $10,000 comes from the $5,000-$10,000 group when λ decreases than when α increases.

Econometric Model of Income Distribution

Having described the family income distribution in the Northeast by the gamma density, a number of important policy implications could be drawn from an econometric model which explains variations in the two parameters, α and λ. The relationships between the two parameters and summary measures of income distribution discussed above help one develop a model relating α and λ to socio-economic factors. The formal mathematical relationship between the parameters is also used to help structure the model.

3/ An added advantage is that the Lorenz concentration ratio (L) is a function of α: $L = 2\beta_{.5}(\alpha,\alpha+1) - 1$; where $\beta_{.5}(\alpha,\alpha+1)$ is the incomplete beta function (Salem and Mount). The coefficient of skewness $1/\sqrt{\alpha}$ varies directly with L.

Since the relationship $\hat{\alpha}/\hat{\lambda} = \overline{X}$ must hold, it is reasonable to include in the structural model an equation for average income. Average income is assumed to be a function of labor productivity (X_1), labor force utilization (X_2 and X_4), factor shares (X_3), and education (X_5).

$$(3) \quad \hat{\alpha}/\hat{\lambda} = a_1 X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} X_4^{\beta_4} X_5^{\beta_5} e_1 .$$

The variables are defined explicitly in table 2; e_1 is a stochastic residual.

In constructing equation (3), it is argued that average family income depends on average productivity capabilities of a region. The second equation relates socio-economic factors to the skewness parameter α. Socio-economic factors which lead to differential income earnings ability of an area's residents are most important. Skewness in income distribution is hypothesized to be a function of minority population concentration (X_9), unequal educational attainment (X_6), labor market characteristics (X_7, X_{10}, X_{11}, X_{12}, and X_{13}), and rural-urban orientation (X_8).

$$(4) \quad \hat{\alpha} = a_2 X_6^{\beta_6} X_7^{\beta_7} X_8^{\beta_8} X_9^{\beta_9} X_{10}^{\beta_{10}} X_{11}^{\beta_{11}} X_{12}^{\beta_{12}} X_{13}^{\beta_{13}} e_2 .$$

(The variables are defined in table 2; e_2 is a random disturbance).

The R^2's are not as high as one would hope, but the large F-statistics and high t-ratios on a number of variables are encouraging. Since $\hat{\alpha}$ appears as part of the dependent variable in both equations, one obtains reduced form equations (substituting equation (4) into equation (3)) to solve for $\hat{\lambda}$. Thus, $\hat{\alpha}$ is completely determined by equation (4) and $\hat{\lambda}$ decreases (average income rises) as X_1, X_2, X_3 and X_5 rise and X_4 falls.

Despite recent questioning of its value, education plays an important role in explaining average family incomes in the Northeast. Productivity

Table 2. Regressions for Parameters of 1970 Family Income Distribution, 70 Northeast Multi-County Areas

Independent Variables[a]	Equation (3) Dependent Variable = $\ln \hat{\alpha} - \ln(\hat{\lambda} x 10^4)$ Regression Coefficient	t-ratio	Independent Variables[a]	Equation (4) Dependent Variable = $\ln \hat{\alpha}$ Regression Coefficient	t-ratio
X_1 = value added per employee	0.411	4.217	X_6 = % of persons over 25 with < 8 years of school	-0.202	3.197
X_2 = workforce as a % of people in working age	0.178	1.354	X_7 = % of persons working 0-26 weeks a year	-0.132	-1.406
X_3 = wage and salary income as a % of total personal income	0.057	0.353	X_8 = % farm population	0.043	3.085
			X_9 = % non-white population	-0.044	-4.651
X_4 = % of persons working 26-50 weeks a year	-0.045	-0.441	X_{10} = % of workers working in county of residence	0.105	1.351
X_5 = median years of schooling for persons over 25	0.871	5.719	X_{11} = Manufacturing Index[b]	-0.159	-2.551
			X_{12} = % wage and salary employment	1.502	3.483
			X_{13} = % of workers in unions	0.032	0.986
Constant	-12.349		Constant	2.167	
R^2 = 0.528; F(5,64) = 14.301			R^2 = 0.653; F(8,61) = 14.354		

Note: Data were obtained primarily from the 1967 Censuses of Business and Manufacturing and the 1970 Census of Population (U.S. Dept. of Commerce, 1970, 1971, 1973). See Boisvert for details.

a/ All variables are in natural logarithms. Equations were estimated by ordinary least squares. No attempt was made to correct for possible correlation between error terms.

b/ Modifying procedures developed by Thurow, 1970, $X_6 = \sum_{i=1}^{n} E_i w_i$ where E_i = % of area's manufacturing employees in the ith manufacturing industry and w_i = ratio of area's average wage industry i to U.S. average wage industry i.

of labor (value added per employee) is also important but its coefficient is less than half as large as the one on education.

The impact of X_1 must be interpreted with caution. Since value added figures were not available for each multi-county area, productivity was estimated on the basis of national value added estimates, weighted by the regional proportion of employment in various sectors. Thus, the variable reflects changes in employment and general levels of productivity and not differences in productivity within any given employment sector across multi-county regions.

Although increased labor participation (a rise in X_2 or fall in X_4) and increased factor shares (X_3) lead to higher average incomes as expected, one hesitates to place much confidence in them because of the low t-ratios. Part of the problem is the inability of X_2 and X_4 to measure underemployment and short-term unemployment of those in the labor force. Wage and salary income as a percent of total personal income was used as a proxy for the portion of the value of regional production going to labor. It did account for the property and profit type income of residents of the area, but payments to factors (other than labor) owned by non-residents could not be isolated. Therefore, to establish the relationship between the functional and size distribution of income, one would certainly need a better measure of factor shares.

Skewness is inversely related to the parameter $\hat{\alpha}$. An increase in $\hat{\alpha}$ results in a more equal distribution of income, a substantial reduction in the portion of families in low income classes and an increase in average income. The impact of education, rural-urban orientation and racial composition on $\hat{\alpha}$ are the easiest to explain. The negative coefficient on X_6

supports the hypothesis that a skewed distribution of educational attainment gives rise to skewed income distribution (Mincer). As the proportion of persons with less than an 8th grade education falls, average income rises and becomes more equally distributed.[4/] The positive coefficient on percent farm population is consistent with the lower average incomes and higher skewness in rural areas than in urban areas. While employment opportunities and composition, as seen below, are important in explaining income differences, the high t-ratio on this variable (X_8) indicates that there are other rural-urban differences which account for these disparities. Finally, in an earlier study Boisvert found that the income disadvantages of areas with large non-white populations declined from 1960 to 1970, but certainly have not disappeared. Results in this study certainly bear out this finding.[5/]

The remaining variables in equation (4) represent different characteristics of the labor market. The importance of wage and salary employment as a percent of all employment (X_{12}) is surprising, even though it was used to account for the impact of factor shares. The fact that $\hat{\alpha}$ increases (i.e., average income rises and skewness falls) as X_{12} increases also reflects the relatively low incomes of farms and other small proprietorships in rural areas.

Although X_{13} has the expected sign, its impact on $\hat{\alpha}$ is smaller than anticipated. Much of its impact could be picked up by X_{12} except the

---

4/ X_6 was used in place of a true measure of skewness because the correlation between the two is high; some skewness values were negative and negative values could not be used in the log-linear functions.

5/ Increased non-white populations lead to only small increases in skewness in the present model, but because of the high t-ratios, discrimination and other factors associated with racial minorities cannot be ignored.

correlation between the two variables is only 0.15. A more plausible reason for the unreliability of this coefficient is the lack of variability across the sample. With the exception of a few multi-county areas which include one or two counties from states south of Virginia and West Virginia, there is little variation in union employment percentages.

The negative coefficient on the manufacturing index (X_{11}) is unexpected since it was constructed to capture the impact of high manufacturing wages. Instead, it seems to reflect the lower incomes and lack of employment opportunities in many old industrialized multi-county regions.

The t-ratios on the remaining two variables are quite low, but it is interesting to speculate on the forces that X_7 and X_{10} measure. The percentage of persons working 0 to 26 weeks a year is intended to measure the skewness of employment effort in much the same way that X_6 measures the skewness of educational attainment. Therefore, reducing the percentage of people working less than 26 weeks a year increases $\hat{\alpha}$ (i.e., increases average income and reduces income inequality). Increasing variable X_{10}, the percentage of workers working in the county of residence, also increases $\hat{\alpha}$, thereby increasing average income, reducing inequality and the proportion of families in low income categories. It appears that efforts to decentralize industry in the Northeast and bring jobs closer to people's places of residence are likely to have some desirable distributional consequences.

Summary and Conclusions

The purpose of this paper has been to demonstrate a convenient method for summarizing the income distribution patterns of a region or group of regions. The advantage of using a continuous density, such as the gamma density, is that one can examine the changes in a number of the income distribution's characteristics simultaneously. This ability is in stark contrast to the limited information generated from an analysis of a summary measure of inequality or the proportion of families in low income categories.

In applying this procedure to 1970 family income distribution patterns in multi-county areas in the Northeast, average income was found to be more responsive to changes in the position parameter of the gamma distribution, λ, while income inequality and the proportion of people in extremely low income categories were more responsive to changes in the skewness parameter, α. Therefore, depending on one's objectives, either to increase the general level of family income or to decrease the income inequality, one should focus on policies to affect λ and α, respectively.

Through econometric analysis, a large proportion of the differences in income distribution among the multi-county regions was explained by the socio-econometric characteristics of the regions. Policies designed to increase average family incomes would be most effective if they focused on raising the average level of educational attainment and the average productivity of workers. Productivity increases could be in the form of increased efficiency of labor in existing employment sectors. However, average productivity could also be increased by encouraging the expansion of those industries in which productivity is already quite high. The anticipated effect on average family income of reducing unemployment and underemployment rates was somewhat smaller, perhaps because many of these individuals have lower productivities. Without education or training and good productive job opportunities, the impact of finding them employment would do little to raise average family income.

If one is more concerned about the equality of income distribution and reducing the number of families with very low incomes, increasing educational opportunities (specifically for those with little or no education) should pay high dividends. The results indicate that a continued removal of the barriers to finding good jobs associated with racial minorities would go hand in hand with improved educational opportunities.

Because incomes tend to be more evenly distributed in areas where a large proportion of the population works in residence (i.e., does not commute), the decentralization of industry throughout the Northeast may have the effect of decreasing income disparities. The substitution of high wage and salary jobs in less populated areas for the lower paying farm and small business alternatives through this decentralization process would affect the sharpest impact on income inequality. But, as in the case of improving average incomes, reduction in unemployment and underemployment would have a small impact unless educational and job opportunities were improved simultaneously.

References

Boisvert, R. N. The Northeast's Changing Low Income Population: An Explanation and Policy Implications. Dep. Agr. Econ. A.E. Res. 76-5, Cornell University, April 1976.

Davis, H. T. Tables of Higher Mathematical Functions, Bloomington, Ind.: Principia Press, Inc., Vol. I, 1933.

Edwards, Clark and Robert Coltrane. "Economic and Social Indicators of Rural Development from an Economic Viewpoint." Southern Journal of Agricultural Economics, No. 1 (1972), pp. 229-245.

Metcalf, C. E. "The Size Distribution of Personal Income During the Business Cycle." American Economic Review, 59 (1969) :657-669.

Mincer, Jacob. "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach." Journal of Economic Literature, No. 8 (1970) :1-27.

Pearson, K. Tables of the Incomplete Gamma Function. London: His Majesty's Stationery Office, 1922.

Salem, A. B. Z. and T. D. Mount. "A Convenient Descriptive Model of Income Distribution: The Gamma Density," Econometrica, 42 (1974) :1115-1127.

Schultz, T. W. "Conflicts Over Changes in Scarcity: An Economic Approach." American Journal of Agricultural Economics, 56 (1974) :998-1004.

Thurow, Lester C. "The Cause of Poverty." Quarterly Journal of Economics, 81 (1967) :40-57.

_______________. "Analyzing the American Income Distribution," American Economic Review, no. 2 (1970), pp. 261-269.

U.S. Department of Commerce, Bureau of the Census. Census of Business, 1967, Washington, D.C., 1970.

_______________. Census of Manufactures, 1967, Washington, D.C., 1971.

_______________. Census of Population, 1970, Characteristics of the Population, Washington, D.C., 1973.