Patterns of Income Concentration Over Time in Puerto Rico: A Curvilinear Regression Model

ARTHUR J. MANN* and WILLIAM C. OCASIO**

Introduction

The relationship between economic growth and the size distribution of income in developing areas has been the focus of debate ever since the original Kuznets presentation.\(^1\) Taking its cue from this seminal contribution, the conventional wisdom appears to posit an inverted U-shaped relation between economic growth (and development) and a given measure of income inequality, with the latter variable increasing in the initial stages of the growth process.

Numerous cross-sectional studies of this relationship have been carried out, employing as a data base both developed and developing countries.\(^2\) Yet, interpretation of the results of such studies is too often clouded by the highly varying quality of the data, especially those of the developing countries. Moreover, quantitative international comparisons that are cross-sectional in nature may give rise to policy inferences or interpretations which are

\(*\)Associate Professor of Economics at the University of Puerto Rico in Mayaguez.

**Graduate student in the Department of Economics at the Massachusetts Institute of Technology.

The authors wish to acknowledge the comments of an anonymous referee.


'For one of the most recent efforts see Montek S. Ahluwalia, "Income Distribution and Development: Somestylized Facts," American Economic Review, V.66 (May, 1976), pp. 120-135.

'directed at specific countries. Kuznets himself has critically questioned such exercises that draw inferences about changes from data on differences.\(^3\) Time-series analysis for developing countries has proved even more treacherous, again in part due to data limitations.

It is the purpose of this paper to use a readily available and consistent data base to analyze the growth-income concentration question over time for one developing area. Both cross-section and time-series analysis will be carried out on data pertinent to the Puerto Rican experience over the 1949-1969 time period. A quadratic relationship between the logarithms of eight socioeconomic independent variables and the Gini coefficient as the dependent variable will be postulated. Thus, since the Kuznets thesis in essence specifies a polynomial function, this paper represents a test of the Kuznets hypothesis with reference to the Puerto Rican case. Such a specification will make it possible to test the inverted U-shaped hypothesis for a specific developing area, a hypothesis that at present appears to be based almost solely upon cross-sectional data covering numerous regions. No attempt is made in this paper to prove or disprove the overall Kuznets thesis. Indeed, it is highly questionable whether or not this can be done with respect to Puerto Rico, given its

unique socio-economic and political relationship with the United States.

Between 1949 and 1969 Puerto Rican real per capita GNP rose from $373 to $994 (in terms of 1954 dollars), or at a rate of close to 5 percent per annum. These high sustained growth rates were accompanied by diverse and profound changes in the island’s socio-economic structure. The model that follows attempts to incorporate some of these changes by employing cross-sectional data pertinent to socio-economic phenomena from the 76 insular municipalities. The data are taken from the census years 1950, 1960, and 1970. Unfortunately, the three years are not completely comparable in that the income-receiving unit is the individual in 1949 and 1959, but is the family in 1959 and 1969 (the 1960 census offers figures for both recipient units). It is for this reason that the subsequent pooling technique and Figures 1A and 1B compare the 1949 and 1959 income distributions of persons and the 1959 and 1969 income distributions of families.

The Explanatory Variables

The explanatory variables listed in Table 1 have been used to quantify the impact of structural changes (which are identified with the economic growth and development processes) on the Puerto Rican income distribution. They were selected on the basis of previous empirical research and a knowledge of Puerto Rican socio-economic conditions. The dependent variable is the Gini coefficient of income concentration (G), where 1 ≥ G ≥ 0. Table 1 summarizes the concept we want to measure, the proxy variable utilized to give operational content to the concept, and the hypothesized sign of the regression coefficient.

A positive sign is hypothesized on the population density variable (X1). Kuznets demonstrated that urbanization, industrialization, and population shifts away from the agricultural sector went hand-in-hand with economic growth and a narrowing of income inequalities in developed countries. However, previous research on income concentration in Puerto Rico over the 1949–1969 time period revealed that urban incomes were invariably more unequal than rural incomes. Given that over this twenty year time span Puerto Rico’s rural/urban population proportions moved from 60–40 to 42–58, it is hypothesized that the higher the level of urbanization, as measured by population density, the greater will be income inequalities.

Using real per capita income (X2) as a proxy for economic growth and development, following Kuznets thesis with respect to developed countries, there exists a tendency for higher real incomes and greater income equality to go together (after a movement toward greater inequality in the initial stages of growth). Yet, Kuznets was not nearly as sanguine regarding the prospects of developing countries. Nevertheless, research on the Puerto Rican experience between 1949 and 1969 concluded that economic growth was accompanied by significant shifts toward decreased inequality. Therefore, a negative sign is hypothesized on this variable.

It is expected that there exists an inverse relationship between the proportion of the labor force employed in manufacturing (X3) and income inequality. Manufacturing generally implies higher capitalization and value productivity than in other sectors. Combining this phenomenon with the application in Puerto Rico of minimum wage legislation and a degree of unionization, the implication is that the greater the proportion of the labor force in manufacturing the more equal will become the concentration of income. Furthermore, previous research has demonstrated that income equalities within the Puerto Rico manufacturing sector are more substantial than in most other economic sectors. It is usually by means of obtaining a formal education that a person enhances his income with job prospects. Sollows found that in the United States the most highly educated groups tended to have more equal income distributions. In fact, it was the group whose family heads had completed high school that showed the most equal relative income concentration. Given that Puerto Rico's educational system is patterned after that of the United States, it is to be expected that a negative relationship will crop up between the index of inequality and the educational level's proxy variable, the number of median school years completed by persons over 24 years of age (X4).

A positive sign is postulated on the discrimination variable as it is so defined by its proxy, the proportion of the labor force forced by females (X5). Discrimination in labor markets leads to a dual wage scale, thereby creating greater overall income dispersion. There are numerous reasons to expect greater skewness in female labor force incomes: women are less readily unionized than men; they accept or are forced to accept lower-paying jobs; they work more on a part-year basis since they are not generally a household's principal breadwinner. Of course, there are other forms of discrimination (religious, racial) for which a direct relation would also be postulated. For the only year (1949) which the data permitted, we included the proportion of non-whites in each municipality as the (racial) discrimination proxy, but, unlike studies referring to the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Concept</th>
<th>Proxy Variable</th>
<th>Hypothesized Sign of Log Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Level of Urbanization</td>
<td>Population Density</td>
<td>+</td>
</tr>
<tr>
<td>X2</td>
<td>Economic Growth</td>
<td>Real Per Capita Income</td>
<td>-</td>
</tr>
<tr>
<td>X3</td>
<td>Economic Growth</td>
<td>Proportion Labor Force in Manufacturing</td>
<td>-</td>
</tr>
<tr>
<td>X4</td>
<td>Education</td>
<td>Median School Years</td>
<td>-</td>
</tr>
<tr>
<td>X5</td>
<td>Education</td>
<td>Completed, Persons Over 24</td>
<td>+</td>
</tr>
<tr>
<td>X6</td>
<td>Industrial Structure</td>
<td>Proportion Labor Force in Agriculture</td>
<td>+</td>
</tr>
</tbody>
</table>


Ibid., pp. 24-27.

The Model and Its Results

Variables $X_1$, $X_2$, and $X_3$ are specified by employing their log and log-squared terms; the remaining independent variables are specified in terms of log alone. The underlying basis for this specification is the inverted U relationship between economic development and income inequality. Such a relationship is not a causal one but is a "stylized fact," and is normally tested by regressing a measure of income inequality on some measure of economic growth (e.g., per capita income). But other measures of economic development must be used. Urbanization ($X_4$) and the percent of the labor force employed in agriculture ($X_5$) represent such measures.

A high activity rate, here operationally defined by the variable labor force as a proportion of the population ($X_6$), most probably signifies a reduced proportion of total families with no or only one income-earner. This, in turn, would lead to a smaller percentage of families (or individuals) falling into the lower end of the income distribution. Therefore, it seems reasonable to postulate an inverse association between the variable and income inequality.

A direct association is posited between the Gini concentration ratio and the proportion of the labor force employed in agriculture ($X_6$), the second proxy variable relating to Puerto Rico's industrial structure ($X_1$ is the first). The agricultural sector displays a higher income concentration than most of the other industrial sectors on the island. Moreover, farm laborers are concentrated at the bottom of the insular size distribution of income, as they tend to be among some of the most unskilled members of the labor force.

*Audic and Mann, op. cit., pp.24-27.
The pooling of years 1949 with 1959 (for persons) and 1959 and 1969 (for families) permits us to combine cross-sectional and time-series analysis. This fashion the effects of variations within years and between years can be observed. In this sense this technique has an advantage over covariance analysis, which merely shows how parameter levels change across years.

With reference to the cross-section results and the coefficient of determination, the eight socio-economic factors accounted for a relatively low 46 percent of the municipality-to-municipality variation in income concentration in 1959 (families), but accounted for a decidedly higher 76 percent of the variation among 1969 families. For 1949 and 1959 persons just over half of the variation was “explained” by the variables. The pooling technique accounted for rather low proportions of overall income variation (under one-third for 1949–1959 persons and a little over one-half for 1959–1969 families).14

Before turning to a discussion of which hypotheses are borne out by the signs on the coefficients in Table 2, let us use the results to explore the inverted U-shaped hypothesis. Has the behavior of income inequality tended to follow an inverted U-shaped pattern in a developing area that has undergone high economic growth rates? In other words, has inequality at first tended to increase and later decrease with growth (and development)?

Figures 1A and 1B present evidence regarding the Puerto Rican growth-inequality issue. It is apparent that what is generally witnessed is not an inverted U-shaped pattern but a variant of an open-ended U-shaped relation. Only in the case of the 1949 data does a distinct inverted U emerge, and even then the downward concavity is not greatly pronounced. For the family income cases (Figure 1B) the right-hand side of the U is quite open and flat in 1969 and in 1959–69 combined, whereas in 1959 the curve for both persons and families is distinctly U-shaped. Thus, initial increases in real per capita incomes are associated with rather rapid decreases in income inequalities.

Therefore, the model yields the general result that the combined effect of the log and log-squared terms of real per capita income ($X_1$) on income inequality is inverse across most of the range of data; i.e., the hypothesized inverse relationship between this variable and the Gini coefficient is borne out. Economic development in Puerto Rico over the twenty year time span covered by this study was associated with greater income equality. It will be noted that $X_1$ is statistically significant at the one percent level for 1959 persons and 1969 families, and at the five percent level for 1949–1959 persons.

This U-shaped relationship between real per capita income and income inequality naturally reflects the net effect on inequality of a host of other variables. The impact of some of these “other” factors is partially quantified by the remaining explanatory variables in the regression. Table 2 reveals that all those remaining variables, with the exception of $X_n$, generally conform to the expected signs on the regression coefficients; i.e., they have the hypothesized signs in those equations in which they are significant. In general terms, those variables which reflect changes in the demographic structure ($X_5$) and the industrial structure ($X_1$) and $X_2$, in labor mobility ($X_6$), and in the economically active portion of the population ($X_7$) are significantly related to inequality. Surprisingly, perhaps, the educational variable ($X_8$) is not statistically significant (except at the five percent level for 1959–1969 families).

The combined effect on income inequality of the log and log-squared terms of variable

14All these results are decidedly “better” (in terms of $R^2$) than those derived from a strictly linear specification using the same variables.
X is positive (in those cases in which it is significant); i.e., the greater the population density (and urbanization) the more unequal the income distribution. In a sense, this would appear to be reasonable in that high population growth rates increase the supply of low-priced labor, thereby maintaining large income differentials.

The negative relation between the proportion of the labor force employed in manufacturing \((X_m)\) and income inequality is not at all surprising. What is interesting is that, under the cross-section results, this relationship is statistically significant only in the case of families, and not for persons. Perhaps this reflects the income-equalizing tendencies of having two (or more) persons in the same family employed. Certainly many women in the Puerto Rican labor force have found employment in manufacturing establishments (especially in light industry, where they are often preferred over males), and the number of families with two or more income-earning members decidedly increased over the 1949-1969 period.

Although variables \(X_2\) and \(X_3\) do conform to the hypothesized relationship, they are marginal in their explanatory power. However, in conjunction with a different group of variables (tested in other earlier models) they do tend to be more statistically important. That variable \(X_3\) is significant in both 1949 persons and 1969 families is noteworthy, for this variable seems to have been little used in other studies. It does point out the importance of the association between labor mobility and income inequality.

The consistent significance (1949 the exception) of variable \(X_4\) serves to emphasize the key role of population involvement in productive activities as a factor associated with income inequality. Finally, with respect to variable \(X_5\), it can be noted that the combined effect of the log and log-squared terms on income inequality is negative. This is rather puzzling, for what it seems to mean is that migration out of the agricultural sector has not contributed toward decreasing income inequality. Perhaps one explanation is that those who leave the sector, generally among the labor force's least skilled members, either fall into the ranks of the unemployed or are forced to accept even lower paying (and part-time) jobs.

### Summary and Conclusion

By using multiple regression techniques and a consistent body of data pertaining to one developing region over a two decade period, we have attempted to analyze statistically the relationship between income concentration and a number of socio-economic variables. The relationship which we most focused upon was that between income concentration and economic growth, as measured by its proxy: real per capita income. We found that a curvilinear regression model in logarithmic form best accounts for Puerto Rican inter-municipal income variations among persons and families over the 1949-1969 time span. This particular specification of an income distribution model describes a U-shaped relation between income inequality and real per capita income. Thus, at least in the case of one developing area, the expected inverted U-shaped pattern does not emerge.

Free enterprise economies have experienced since World War II a continued rise in the general level of prices, downward rigidity of wages, and, since the late sixties, the strange phenomenon of a rising general price level, accompanied by high and rising levels of unemployment during recessionary phases of the business cycle. These features of modern free enterprise economies are but reflections of what may be described as changes in the outer-level parameters of modern economic systems, and have their sources mainly in the evolution and maturity of big business, big labor, and big government. All three institutions exercise considerable market power, and contribute thereby to the twin problems of inflation and unemployment.

Associated with the development of the three institutions are reciprocal changes in attitudes, expectations, and behavior of the ordinary man in the street operating as a consumer, worker, or investor. For example, the behavioral assumption that the wage earner will make decisions based on changes in his money wage rather than his real wage, has lost much of its validity. The working man as well as the businessman, has come to expect a continuing rise in the general level of prices based on his past experience, and his behavior when renewing contracts, or determining price of output, is designed to protect himself from the anticipated reduction in income. Hence the widening application of escalator clauses to transaction and even to income-transfer arrangements. This behavior has led to inflationary expectations becoming embedded in the structure of the U.S. economy and other modern economies and complicates the problem of prescribing policy for dealing with high unemployment. There is now widespread belief in modern societies that fiscal and/or monetary policy aimed at reducing unemployment will be inflationary, and that attempts to provide full employment would trigger galloping inflation. A critical aspect of the fight against inflation is that of arresting and reversing the widespread expectation of continuing inflation.

The continued rise in the general level of prices is explained as due to demand-pull and cost-push forces. Both forces might be at work interacting with and reinforcing each other. In its pure form, demand-pull inflation exists when there is excess aggregate demand in any economy operating at or close to the full-employment level. This is the familiar brand of inflation for controlling which fiscal and monetary policy instruments have been devised. Cost-push inflation is the stranger which is not susceptible to control by traditional anti-inflationary policy. It persists in the face of high or rising unemployment.

A Proposal for Controlling Stagflation: A New Automatic Stabilizer

H. G. HENRY
Colgate University

---
