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 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 tenuous, 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

*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.

[1]Simon Kuznets, "Economic Growth and Income Inequality," *American Economic Review*, V.45 (March, 1955), pp. 1–28.

[2]For one of the most recent efforts see Montek S. Ahluwalia, "Income Distribution and Development: Some Stylized Facts," *American Economic Review*, V.66 (May, 1976), pp. 128–135.

[3]Simon Kuznets, *Modern Economic Growth* (New Haven: Yale University Press, 1966), pp. 433–437.

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.[4] 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.

[4]U.S. Dept. of Commerce, Bureau of the Census, *U.S. Census of Population: 1950, 1960,* and *1970, Puerto Rico, General Social and Economic Characteristics* and *Detailed Characteristics* (Washington: U.S. Govt. Printing Office, 1952, 1962, and 1972 respectively). The values of all variables except income actually refer to the years 1950, 1960, and 1970, whereas the income variable covers the years 1949, 1959, and 1969.

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 $1 \geq G \geq 0$. 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 (X_1). Kuznets demonstrated that urbanization, industrialization,

Table 1 Explanatory Variables Used for the Regression Analysis

Variable	*Concept*	*Proxy Variable*	*Hypothesized Sign of Log Variable*
X_1	Level of Urbanization	Population Density	+
X_2	Economic Growth	Real Per Capita Income	−
X_3	Industrial Structure	Proportion Labor Force in Manufacturing	−
X_4	Educational Level	Median School Years Completed, Persons Over 24	−
X_5	Discrimination	Females as Proportion of Labor Force	+
X_6	Labor Immobility	Proportion of Population Born in Municipality of Residence	+
X_7	Activity Rate	Labor Force as Proportion of Population	−
X_8	Industrial Structure	Proportion Labor Force in Agriculture	+

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.[5] 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 (X_2) as a proxy for economic growth and development, following the 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.[6] 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 (X_3) 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 Rican manufacturing sector are more substantial than in most other economic sectors.[7]

It is usually by means of obtaining a formal education that a person enhances his income with job prospects. Soltow[8] found that in the United States the more 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 (X_4).

A positive sign is postulated on the discrimination variable as it is so defined by its proxy, the proportion of the labor force formed by females (X_5). 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 posited. 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

[5]See Fuat M. Andic and Arthur J. Mann, "Secular Tendencies in the Inequality of Earnings in Puerto Rico," *Review of Social Economy,* V.34 (April, 1976), pp. 13–32.

[6]Ibid.

[7]Ibid., pp. 24–27.

[8]Lee Soltow, "The Distribution of Income Related to Changes in the Distribution of Education, Age, and Occupation," *The Review of Economics and Statistics,* V.42 (November, 1960), pp. 450–453.

United States, it did not turn out to be statistically significant.

Greater labor market mobility would seem to improve one's job and income opportunities and prevent the emergence of monopsonistic markets. The greater the labor supply inelasticity (and immobility), the more depressed become wage rates as demand diminishes. The large Puerto Rican emigration to the United States may have acted to decrease income inequalities on the island by removing from the labor force the unskilled and lesser-skilled, who are to found concentrated at the bottom of the income distribution. Thus, *a priori* there would appear to exist a positive relationship between labor immobility as measured by its proxy—the proportion of the population born in its municipality of residence (X_6)—and income inequality.

A high activity rate, here operationally defined by the variable labor force as a proportion of the population (X_7), 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 this 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_8), the second proxy variable relating to Puerto Rico's industrial structure (X_3 is the first). The agricultural sector displays a higher income concentration than most of the other industrial sectors on the island.[9] 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.

[9]Andic and Mann, op.cit., pp.24–27.

The Model and Its Results

Variables X_1, X_2, and X_8 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_1) and the percent of the labor force employed in agriculture (X_8) represent such measures.

On the other hand, variables X_4 through X_7 are factors related to the labor force, and there is no *a priori* reason why an inverted U or U relationship between them and the measure of inequality is to be expected. Human capital theory certainly leads one to expect an inverse relationship between median years of education and income inequality, and there seems to be no reason why this relationship should change direction as the median increases. A similar argument can be constructed for X_5 and X_6, which may be interpreted as proxies for labor market imperfections. With respect to X_7, the activity rate, one would always expect an inverse relationship to exist. A major problem regarding X_3 does arise, however. In this instance *a priori* expectations would generate an inverse association with income inequality. But an inverted U-shaped relationship may be at least plausible. However, as soon as variables are squared the problem of multicollinearity becomes critical, even though cross-sectional data are used. If the *a priori* specification of the model were based on sound economic theory, and if the theory stated that an inverted U relation was called for, then the square of the variables would have been included, since the Gauss-Markov theorem states that ordinary least squares (OLS) yields the best linear unbiased estimates (BLUE), and omission would introduce a bias in all estimates. However, since a strong theoretical foundation is lacking, and since the model is not strictly causal, we do not square X_3 to reduce multicollinearity.

Table 2 presents the results of the estimating equations. The cross-section results appear in the first four columns of the table, while the pooled results are presented in the

Table 2 Cross-Municipality Non-Linear Regressions Explaining Income Inequality, Puerto Rico, 1949, 1959, and 1969 (Standard Errors in Parentheses)

	Cross-Section Results				*Pooled Results*	
Explanatory Variables	*1949 Persons*	*1959 Persons*	*1959 Families*	*1969 Families*	*1949–59 Persons*	*1959–69 Families*
Constant	798	4381	2442	3662	1512	1025
$\log X_1$	268.00**	126.69*	108.28*	−49.44	142.20**	−49.84
	(88.80)	(57.81)	(52.16)	(44.93)	(54.29)	(36.07)
$[\log X_1]^2$	−19.11**	−10.22*	−8.82*	3.40	−10.62**	3.52
	(6.71)	(4.28)	(3.86)	(3.30)	(4.04)	(2.65)
$\log X_2$	−566.68	−1160.19**	−462.34	−844.96**	−351.40*	30.44
	(358.63)	(302.51)	(272.98)	(291.95)	(174.86)	(126.68)
$[\log X_2]^2$	56.59	106.58**	43.38	71.09**	32.93*	−.78
	(33.83)	(26.45)	(23.87)	(22.79)	(15.75)	(10.74)
$\log X_3$	−11.90	−12.90	−14.37*	−22.89**	−20.47**	−20.21**
	(8.33)	(7.84)	(7.07)	(7.69)	(6.19)	(5.58)
$\log X_4$	.00	−28.40	−11.66	−22.44	−16.58	−35.93*
	(17.28)	(18.53)	(16.72)	(25.80)	(14.16)	(15.02)
$\log X_5$	47.59*	32.91	34.77	21.75	63.68**	18.49
	(19.60)	(20.22)	(18.25)	(24.62)	(15.65)	(15.03)
$\log X_6$	79.36*	58.20	6.59	110.75**	21.47	29.59
	(33.91)	(34.79)	(31.39)	(24.81)	(24.81)	(20.55)
$\log X_7$	−83.75	−223.59**	−160.58**	−125.25**	−114.22**	−79.09**
	(52.04)	(58.09)	(52.42)	(31.21)	(33.18)	(18.44)
$\log X_8$	17.89	−159.19**	−163.74**	−157.95**	−95.89**	−106.06**
	(48.60)	(42.99)	(38.80)	(46.95)	(35.91)	(28.97)
$[\log X_8]^2$	−.81	11.50**	12.63**	12.75**	7.50**	9.29**
	(3.34)	(3.08)	(2.78)	(3.83)	(2.51)	(2.25)
$\bar{R}^2$	.54	.52	.46	.76	.27	.52
F-Ratio	6.70**	6.28**	4.94**	18.03**	4.83**	13.96**

*Significant at 5 percent level.
**Significant at 1 percent level.
X_1 = Population Density
X_2 = Real Per Capita Income
X_3 = Proportion Labor Force Employed in Manufacturing
X_4 = Median School Years Completed, Persons 25 years and older
X_5 = Female Proportion of Labor Force
X_6 = Proportion of Population Born in Municipality of Residence
X_7 = Activity Rate
X_8 = Proportion Labor Force Employed in Agriculture

last two columns. 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. In 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 haw 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).[10]

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)?

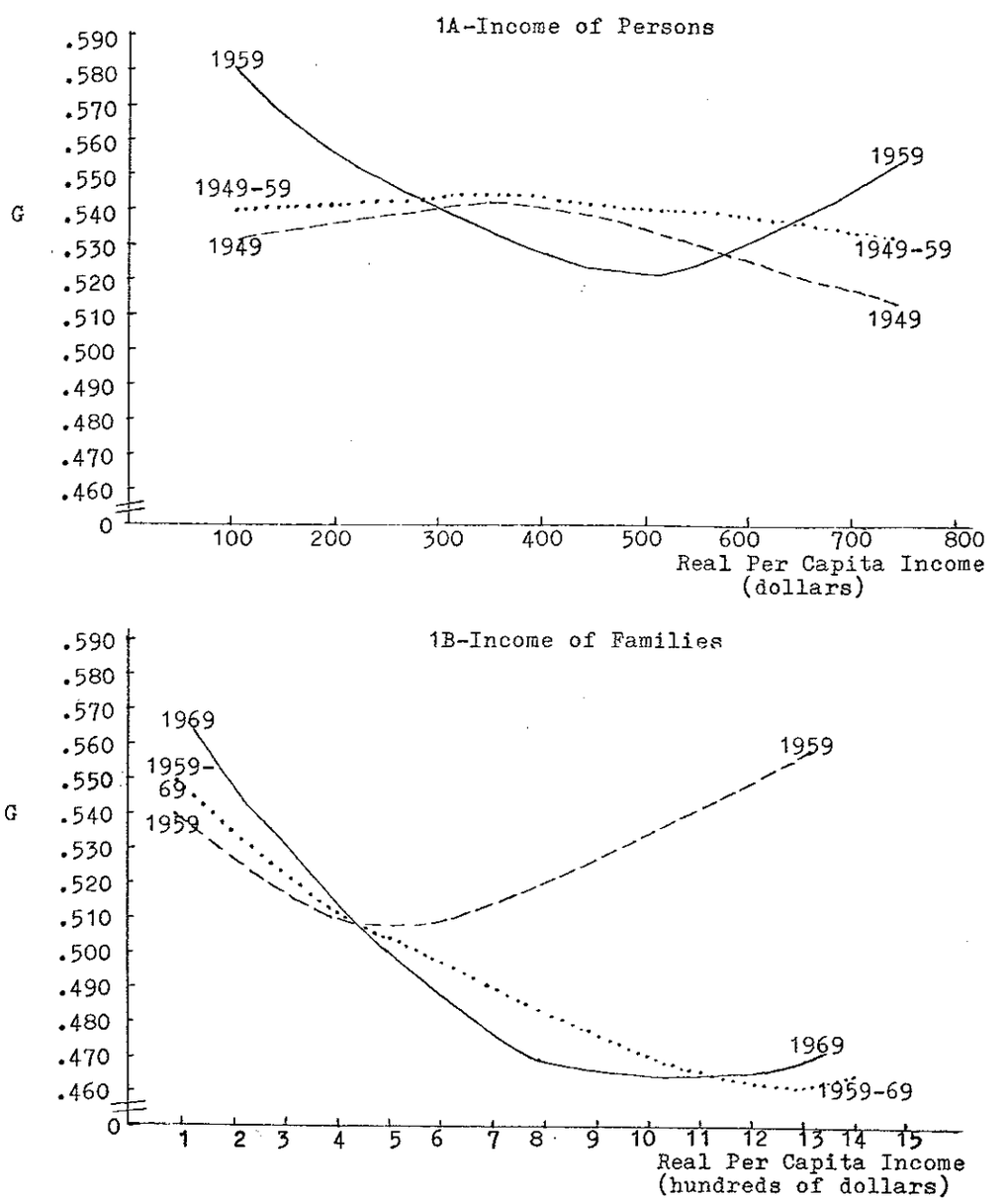
Figures 1A and 1B present evidence regarding the Puerto Rican growth-inequality issue. It is apparent that what is generally witnessed is not an inverted U-shaped pattern but a variant of an open-ended U–shaped relation. Only in the case of the 1949 data does a distinct inverted U emerge, and even then 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_2) 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_2 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 these remaining variables, with the exception of X_8, 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_1) and the industrial structure (X_3) and X_8), in labor mobility (X_6), and in the economically active proportion of the population (X_7) are significantly related to inequality. Surprisingly, perhaps, the educational variable (X_4) 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

[10]All these results are decidedly "better" (in terms of $\bar{R}^2$) than those derived from a strictly linear specification using the same variables.



Figures 1A and 1B Estimated Relationships Between Real Per Capita Income and Gini Coefficient (G) of Income Inequality, Puerto Rico, 1949–1969

X_1 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_3) 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_4 and X_5 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_6 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_7 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_8, 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.