Fields, Gary S.

Working Paper

Income Inequality in Urban Colombia: A Decomposition Analysis

Center Discussion Paper, No. 267

Provided in Cooperation with:
Yale University, Economic Growth Center (EGC)

Suggested Citation: Fields, Gary S. (1977) : Income Inequality in Urban Colombia: A Decomposition Analysis, Center Discussion Paper, No. 267, Yale University, Economic Growth Center, New Haven, CT

This Version is available at:
https://hdl.handle.net/10419/160194

ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station
New Haven, Connecticut

CENTER DISCUSSION PAPER NO. 267

INCOME INEQUALITY IN URBAN COLOMBIA: A DECOMPOSITION ANALYSIS

Gary S. Fields

September 1977

The research for this paper was conducted at the Centro de Estudios sobre Desarrollo Económico, Universidad de Los Andes, Bogotá, Colombia, and at the Economic Growth Center, Yale University. Partial support for this research was received from the International Bank for Reconstruction and Development under RPO/284. However, the views expressed do not necessarily reflect those of IBRD. The author wishes to thank the above institutions without implicating them. He also acknowledges the valuable research assistance of Helena de Jaramillo, Judith Oder, and David Bruce as well as the comments of Gustav Ranis.

INCOME INEQUALITY IN URBAN COLOMBIA: A DECOMPOSITION ANALYSIS

Gary S. Fields
Yale University
August, 1977

Introduction

The persistence of poverty and income inequality in less developed countries (LDCs) is a source of serious concern to development economists. To understand the structure of inequality, several researchers using a variety of methodologies have measured the importance of various contributory factors to overall income variability. The available literature---which now includes studies of Brazil, Mexico, Iran, the Philippines, Taiwan, Thailand, Pakistan, and Colombia---has been reviewed elsewhere (Fields, 1977). This paper presents additional evidence for urban Colombia, in the process raising some important methodological issues which bear on the design of future research studies.

The data set used in this paper is described in Section I. The decomposition of Colombian inequality by functional income source is presented in Section II for micro data. Section III examines the robustness of source decomposition procedures to data aggregation. Section IV presents inequality decompositions by city, and Section V by other income-determining characteristics. Conclusions appear in Section VI.

I. The Data

In late 1967 and early 1968, the Center for the study of Economic Development (CEDE) at the University of the Andes in Bogota, Colombia carried out a family budget study in the four major cities of Colombia.[1] This survey, known by the Spanish acronym PRESFAM, yielded detailed data on the spending patterns, income sources, and family characteristics of 2,949 households. Computer tapes containing the coded questionnaire responses were generously provided by CEDE and by the Program of Joint Studies of Latin American Economic Integration (ECIEL).

For purposes of this paper, the most important aspects of the data set are the income variables and the personal characteristics. Total income refers to the family's income from all sources in the three months preceding the survey and includes income-in-kind and imputed rent. The family's total income is broken down according to income from various sources. Wage income includes wages, salaries, overtime payments, profit-sharing, and value of on-the-job income received in kind. Independent income refers to the net income from independent work in a business, profession, or domestic service. Capital income includes interest, dividends, rents, and imputed rents for owner-occupied housing. Finally, transfer income is defined to include both private and public transfers such as pensions, social benefits, and students' scholarships. Information is available on the following personal characteristics of the head of the household: education, occupation, employment status, sector of the economy, age, and sex. For further information on the PRESFAM data, see Prieto (1971), Musgrove (1974), and Fields and Jaramillo (1975).

---

[1] These cities are Bogota, Barranquilla, Cali, and Medellin. Their respective populations in the most recent preceeding Census were: Bogota, 1,697,300; Medellin, 772,900; Cali, 637,900; Barranquilla, 498,300.

II. Decomposition of Urban Inequality by Functional Income Source: Micro Data

Source decompositions have been carried out in studies of Taiwan by Fei and Ranis (1974) and Fei, Ranis, and Kuo (1977) and of Pakistan by Ayub (1977). The question asked in source decompositions is: of total inequality, how much is attributable to income from wage labor, how much to income from independent labor, how much to income from capital, and how much to income from transfers? The empirical analysis of this section quantifies these effects for urban Colombia and further shows the way in which each source's contribution to overall inequality depends positively on the degree of inequality of each income source, the importance of that income source in total income, and the extent of correlation between income from that source and total income.

The methodology for source decompositions developed by Fei and Ranis uses the Gini coefficient as the measure of inequality. Gini coefficients for total income and for each functional income source are calculated. Also required for each income source is a so-called pseudo-Gini coefficient, i.e., the Gini coefficient that would be obtained for that factor's income if the families were ordered according to total income rank rather than according to their income from that particular income source. It is shown that the overall Gini for total income (G) is a weighted average of the pseudo-Ginis for the i'th income source ($\bar{G}_i$) with the weights given by the factor share of that income source (ϕ_i):

$$(1) \quad G = \bar{G}_1 \phi_1 + \bar{G}_2 \phi_2 + \bar{G}_3 \phi_3 + \bar{G}_4 \phi_4.$$

The pseudo-Gini for the i'th source ($\bar{G}_i$) is equal to the product of the true Gini for that source (G_i) and a relative correlation coefficient (R_i), defined below:

$$(2)\quad \bar{G}_i = G_i R_i.$$

For each factor, the relative correlation coefficient is the ratio of two other correlations:

$$(3)\quad R_i = \frac{\text{cor}(Y_i,\rho)}{\text{cor}(Y_i,\rho_i)} = \frac{\text{coefficient of correlation between factor income amount and total income rank}}{\text{coefficient of correlation between factor income amount and factor income rank}}.$$

To further explain (3), consider the R_i for wage income. The numerator of (3) is the correlation between wage income in dollars (Y_i) and the family's total income position (ρ), ordered from lowest to highest. The denominator of (3) relates the dollar wage income figure (Y_i) to that family's wage income rank (ρ_i).

Substituting (2) and (3) into (1) and dividing through by G, we obtain:

$$(4)\quad 100\% = \phi_1 \frac{G_1}{G}\frac{\text{cor}(Y_1,\rho)}{\text{cor}(Y_1,\rho_1)} + \phi_2 \frac{G_2}{G}\frac{\text{cor}(Y_2,\rho)}{\text{cor}(Y_2,\rho_2)} + \phi_3 \frac{G_3}{G}\frac{\text{cor}(Y_3,\rho)}{\text{cor}(Y_3,\rho_3)}$$
$$+ \phi_4 \frac{G_4}{G}\frac{\text{cor}(Y_4,\rho)}{\text{cor}(Y_4,\rho_4)} = FIW_1 + FIW_2 + FIW_3 + FIW_4,$$

the FIW's denoting the so-called Factor Inequality Weights of wage income, independent labor income, capital income, and transfer income respectively. Equation (4) shows explicitly the dependence of overall inequality on the degree of inequality of each income source, the extent of correlation between income from that source and total income, and the importance of that income source in the total.

Applying this source decomposition methodology to the microeconomic data for urban Colombia at the household level, we obtain the decomposition statistics given in Table 1. The outstanding result is that labor income (wage plus independent) accounts for the bulk of overall income inequality (70%) whereas capital income accounts for 26% of inequality and transfer income for 4%. This finding is at odds with the usual perception that disparities in holdings of wealth are the principal source of inequality in Colombia and elsewhere. An explanation for this result must be sought.

Looking behind the Factor Inequality Weights is revealing. We see from the factor Gini coefficients (G_i) that, as expected, capital income and transfer income are highly unequally distributed and that labor income is distributed much more equally. How then can labor income be accounting for so much of overall inequality? Part of the answer is to be found in the correlational patterns. The correlation between total income and factor income (cor Y_i,ρ) is much greater for labor income than for other income sources. These correlations, though positive, are far from unity, even for labor income. Now, the factor incomes shares also enter in. Not only is labor's functional share so much larger but it is also the case that most families in urban Colombia (84%) receive most if not all of their income from the work they do (see Table 2). Hence, in the majority of cases, high labor income and high total income go hand-in-hand, and similarly for low labor and total incomes. The reason that labor income contributes so much to overall inequality, therefore, is that labor income is so important a part of total income and it is distributed far from equally.

Table 1.

Decomposition of Inequality in Urban Colombia

by Functional Income Source, 1967-68, Based on Microeconomic Data

	Wage Income	Indep. Labor Income	Capital Income	Transfer Income	Total Income
Factor Income Share (ϕ_i)	.3527	.3467 (.6994)	.2186	.0820	1.0000
Gini Coefficient(G_i)	.6830	.8291 (.5692)	.7901	.8297	.5085
Correlation between Factor Income Amount and Total Income Rank (cor Y_i,ρ)	.4183	.4474 (.6574)	.3984	.1653	
Correlation between Factor Income Amount and Factor Income Rank (cor Y_i,ρ_i)	.7334	.6009 (.7446)	.5115	.5013	
Relative Correlation Coefficient (R_i)	.5704	.7445 (.8828)	.7789	.3297	
Pseudo-Gini Coefficient($\bar{G}_i$)	.3896	.6173 (.5025)	.6154	.2736	
Factor Inequality Weight (FIW_i)	.2702	.4208 (.6910)	.2647	.0442	1.0000

Table 2.

Analysis of Income Sources in Urban Colombia,

1967-68, Based on Microeconomic Data

Percentage of Families Having Some Income from Each Source:

Wages and Salaries	63%
Independent Labor Income	40%
Salaries and/or Indep. L	90%
Capital (including imputed rent)	59%
Transfer	46%

Relationship Between Labor Market Income and Other Income:

Total Labor Income = Wage Income + Independent Labor Income

	0	>0	Row Total
Other Income=0	8 (0.3%)	718 (24.4%)	726 (24.6%)
0<Other Income≤Labor Income	0 (0.0%)	1742 (59.1%)	1742 (59.1%)
Other Income>Labor Income	285 (9.7%)	196 (6.7%)	481 (16.3%)
Column Total	293 (10.0%)	2656 (90.0%)	2949 (100.0%)

In sum the decomposition of inequality by functional income source in urban Colombia reveals that more than two-thirds of overall inequality is attributable to labor income. The principal inequality-producing factor is that some people receive a great deal more income for their work than do others. The intuitively-plausible prior notion that the most unequally-distributed factors contribute the most to total inequality is found to be false in this case. In Taiwan, which serves as a prototype for this type of calculation, and in Pakistan, where the data permit such calculations, the preeminence of labor income inequality has also been found.

One significant feature of the computations for Colombia is that all Gini coefficients and correlation ratios are based on individual families, not on family groupings. Past researchers have not had access to such disaggregated data. An interesting question is which, if any, of the findings for Colombia would have been altered if only aggregated data had been available. The results of a parallel decomposition exercise for urban Colombia based on family groupings rather than on individual families are reported in Section III. As we shall see, in some respects, the two sets of results differ substantially.

III. Source Decompositions and Data Aggregation

Often, statistical publications tabulate data in ways different from what researchers interested in particular problems would have specified. This problem is especially acute in less developed countries, where data are so much scarcer. In Colombia, though, we are fortunate to have access to the survey questionnaires for each family. A rare opportunity to perform a controlled experiment arises. By aggregating the data as they have been tabulated elsewhere, we are able to determine which of the Colombian results are robust to grouping of data and which are not. By analogy, results from the Colombian experiment can be used to infer how advisable it is to work with family groups when the choice is between grouped data and nothing.

The aggregated data are presented in Table 3. Following the aggregation procedure used in existing data sources in other countries, families are grouped according to total income. Their incomes from each factor are summed and averaged. Thus, for example, in the 0-100 peso income group, the mean income is 78.3 pesos. Of that 78.3, on average 14.8 is from wage income, 24.2 from independent labor income, and so on.

The decomposition statistics from grouped data are presented in Table 4. When these are compared with those from ungrouped data (Table 1), both similarities and differences emerge. The Gini coefficients themselves differ by less than one percent.[1] Functional income shares are identical, as indeed they should be. Surprisingly, the pseudo-Gini coefficients and hence the factor inequality weights are virtually the same in the two tabulations, the differences being so small as to be ascribable to the use of rank correlation coefficients in one calculation and ordinary

[1] The Gini coefficient for total income computed from micro data is .5085 and from grouped data .4965, the difference between the true and the estimated values being due to the neglect of within-group inequality in the latter.

TABLE 3 AVERAGE TOTAL & COMPONENT INCOME.

1967 - DECOMPOSITION IS INCOME CLASS BY INCOME

COLOMBIA

	GROUP (00)	# OF HSHOLDS	TOTAL INCOME	WAGE INCOME	INDEP INCOME	CAPTL INCOME	TRANS INCOME	MISC INCOME
1	0	3.0000	0.0	0.0	0.0	0.0	0.0	0.0
2	0-1	19.0000	78.2632	14.8421	24.2105	12.0000	12.4737	14.7368
3	1-2	173.0000	158.6994	74.4855	47.9480	14.1387	15.7861	6.3410
4	2-3	324.0000	250.5494	113.2284	69.4722	29.1389	30.2839	8.4259
5	3-4	363.0000	344.7327	153.8788	79.2121	46.9587	49.6722	15.0110
6	4-5	298.0000	444.8120	219.3121	83.1476	62.8859	60.4530	19.0134
7	5-6	236.0000	545.8770	251.5085	101.9025	98.4449	74.4110	19.6102
8	6-7	232.0000	640.0559	252.1164	180.6034	99.7758	82.7069	24.8534
9	7-8	166.0000	747.9397	338.4336	153.0903	131.0723	89.7771	35.5663
10	8-9	122.0000	843.0574	318.3032	201.8852	133.4672	138.7213	50.6803
11	9-10	111.0000	940.2341	365.5854	280.6846	167.7477	97.3243	28.8919
12	10-11	81.0000	1051.2097	396.2715	274.4072	202.9876	142.5802	34.9630
13	11-12	64.0000	1144.9844	440.7031	294.7969	259.3750	97.4219	52.6875
14	12-13	83.0000	1233.8794	441.0601	343.7107	283.6384	116.4699	49.0000
15	13-14	56.0000	1348.4819	472.1428	414.6069	302.6606	80.5536	78.5179
16	14-15	32.0000	1434.6250	587.1562	311.2500	334.8750	119.4687	81.8750
17	15-16	53.0000	1534.9810	626.3960	434.0000	213.6981	185.0189	75.8679
18	16-17	27.0000	1647.0369	745.9258	388.3333	248.2222	148.1111	116.4444
19	17-18	38.0000	1736.2893	754.4736	292.6577	326.9473	110.2105	252.0000
20	18-19	42.0000	1827.0476	634.8333	547.3809	354.3809	244.3810	46.0714
21	19-20	33.0000	1946.6060	850.2119	565.9089	362.6362	46.6970	121.1515
22	20 PLUS	393.0000	3726.6741	884.7302	1557.8777	913.0991	129.0356	241.9313
23	COUNTRY	2949.0000	1070.2375	356.4172	350.3772	220.8691	82.9091	59.6646

Table 4

Decomposition of Inequality in Urban Colombia
by Functional Income Source, 1967-68, Based on Grouped Data

	Wage Income	Indep. Labor Income	Capital Income	Transfer Income	Total Income
Factor Income Share (ϕ_i)	.3527	.3467	.2186	.0820	1.0000
	.6994				
Gini Coefficient (G_i)	.3858	.5951	.5860	.2973	.4965
	.4886				
Correlation between Factor Income Amount and Total Income Rank (cor Y_i,ρ)	Not computed				
Correlation between Factor Income Amount and Factor Income Rank (cor Y_i,ρ_i)	Not computed				
Relative Correlation Coefficient (R_i)*	.9986	.9947	.9985	.9139	
	.9999				
Pseudo-Gini Coefficient ($\bar{G}_i$)	.3854	.5936	.5848	.2688	
	.4886				
Factor Inequality Weight (FIW_i)	.2765	.4187	.2546	.0448	1.0000
	.6952				

*Coefficient of rank correlation

correlations in the other. Where the two sets of calculations diverge is in the breakdown of the factor inequality weights. The factor Ginis estimated from grouped data are a great deal lower than the true values, differing by the following percentages: wage, 77%; independent labor, 39%; capital income, 35%; transfer income, 279%. On the other hand, **in the grouped data, the coefficients of correlation between each factor income amount and total income (.91 to .99) are too high, unbelievably** so. The extent of overstatement is, of course, the same as the degree of understatement of the factor Ginis, the reason being that the product of the two (the pseudo-Gini coefficient) is nearly the same for each income type. Thus, it may be concluded that although the overall Gini coefficients, the factor income shares, the factor inequality weights and pseudo-Gini coefficents are comparable for grouped and ungrouped data, the factor Gini coefficients and correlation ratios obtained from grouped data provide substantially distorted estimates of the true values.

Intuitively, it is not hard to see why the type of grouping in Table 3 leads to such distorted estimates. Recall that the factor incomes reported in any row of the table are the sums for all families in that total income class. Some of those families may have no income from any given factor, other families may receive all their income from that factor, and the rest are scattered in between.[1] The families with zero income from a particular factor are averaged in with families with positive incomes from that factor in the same total income class. For example, if the 0-100 peso income class were comprised of two families, one with 50 pesos of wage income, the other with 50 pesos of capital income,

---

[1] In actuality, the percentages are substantial: 37% with no wage income, 60% with no independent labor income, 41% with no capital income, and 55% with no transfer income.

Table 3 would report a group of two families with average wage income of 25 pesos and average capital income of 25 pesos. Thus, all the zero factor income cases disappear, as do the high factor income cases.[1] The result, not surprisingly, is a large diminution in apparent factor income inequality. Contrarywise, because of all the averaging and the fact that total income is the sum of its parts, the average factor incomes across income classes must increase nearly monotonically almost by definition, except when the factor is a small part of the total. That the coefficients of correlation between factor income and total income groups approach one under such circumstances is both understandable and artifactual, as is the seeming observation in Table 4 that wage and transfer income are distributed more equally than total income and independent and capital income less so. The difficulty with the factor Gini coefficients could have been avoided very simply had the factor income groups been based on the amount of factor income rather than on the amount of total income, but then we would have had no information on the R's.

What do the results of this Section imply about the conduct of decomposition analysis? Our goal is to understand the structure of inequality in a given country at a point in time or changes in inequality over time. The factor inequality weights calculated from grouped data closely approximate the weights calculated from micro data. Thus, if the concern is with assessing the relative importance of income from labor, capital, or transfers in accounting for income inequality and using the resulting information to decide whether to concentrate subsequent research efforts on studies of labor markets, wealth

[1] 35% of the families in the PRESFAM Sample in Colombia received all their income from one source only, yet nowhere in Table 3 are factor incomes and total incomes equal.

holdings, or government tax and transfer schemes, grouped data work fine. But decomposition analysis is often carried further and is used to break down the factor inequality effects in terms of inequality components, i.e., functional income shares, correlations between factor incomes and total income, and factor inequality. The evidence presented above for urban Colombia shows that only the first of these is measured from grouped data with any accuracy. This suggests that for this particular decomposition problem with this particular type of grouped data, the option of doing nothing at all rather than using what imperfect data we have deserves serious consideration.

Let us now turn from the source decomposition problem to other types of inequality analysis.

IV. Decomposition of Urban Inequality by City

Several writers have observed differentials in average incomes and expenditures between one Colombian city and another. Prieto (1971, Part III, Table 1), for instance, reported the following mean family expenditures (in pesos per three months):

Bogota Col.	$8,150
Barranquilla	$7,090
Cali	$6,640
Medellin	$5,980
Average four cities	$7,230

Isaza and Ortega (1971) found similar differences. Because of these differentials, Musgrove (1974) analyzed incomes in each Colombian city separately. Berry and Urrutia's recent book (1976) devoted a chapter to exploring interregional and intercity inequality. Many other examples could undoubtedly be adduced in the Colombian context. Elsewhere, the works of Kuznets (1963) and Williamson (1965) on interregional inequality stand out.

In light of these concerns, it is interesting to ask how much income variability in Colombia is associated with differences across the various cities and how much to differences within them. A number of methodologies are available for addressing this question. A particularly comprehensive statistical procedure, and the one used here, is analysis of variance (ANOVA).

In our problem, the dependent variable is the logarithm of family income in each of the nearly 3,000 sample households and the independent variable is the city of residence. The variance, which is the sum of squared deviations from the mean (SS), is expressed as:

$$(5) \quad SS_y = SS_{\text{between cities}} + SS_{\text{within cities}}$$

where $SS_y = \Sigma\Sigma_{ji} (Y_{ji} - \bar{Y})^2$ in which $\bar{Y}$ is the overall mean of log income Y in the entire sample, the i's are households, and the j's are various cities

$SS_{\text{between cities}} = \Sigma_j N_j(\bar{Y}_j. - \bar{Y})^2$ in which $\bar{Y}_j.$ is the mean income in city j, and N_j is the number of sample households in city j

and $SS_{\text{within cities}} = \Sigma\Sigma_{ji} (Y_{ji} - \bar{Y}_j.)^2$

In this way, equation (5) tells us the relative importance of income inequality within cities as compared with diversity in mean incomes across cities. Additionally, and quite importantly, tests of statistical significance are available for each factor.

The ANOVA results for the city decomposition are reported in Table 5. City is significant statistically but not economically in explaining urban inequality. Given the large size of the sample, the income differences observed across Colombian cities are found to be significant statistically, the F ratio of 3.825 surpassing the .01 significance level. Nonetheless, a negligible share of the variance in log income---only 0.4%---is explained by variation <u>across</u> cities. Nearly all of the inequality in urban Colombia is due to variations <u>within</u> cities. Despite the intercity wage differentials stressed by some authors, knowledge of a family's city of residence provides very little information on its income.

Can we get further with other family information? This question is explored in Section V.

V. Decomposition of Urban Inequality by Income Determinants

This Section presents the results of analysis of variance (ANOVA) by income determinants.[1] To look further for explanations of incomes and to account for income inequality, the findings of Section II suggest the usefulness of close examination of labor income inequality. It is known that labor earnings in Colombia are related systematically to characteristics of workers, characteristics of employers, and characteristics of industries.[2] Let us now consider two variables which receive frequent mention --- education and age --- along with city of residence.

[See next page for footnotes]

Table 5.

Decomposition of Inequality in Urban Colombia by City, 1967-68

Dependent Variable: Log Variance

Source of Variation	Sum of Squares		F	Significance of F
Main Effect Explained by City	9.8	(0.4%)	3.825	.01
Unexplained	2519.4	(99.6%)		
Total	2529.3	(100.0%)		

ANOVA can handle multiple explanatory variables, breaking down the log variance of income in the following way:

(6) SS_y = SS due to city + SS due to education + SS due to age
+ SS due to city-education interactions
+ SS due to city-age interactions
+ SS due to education-age interactions
+ SS due to city-education-age interactions
+ SS within city-education-age groupings

From a decomposition like (6), we can learn: whether income inequality is greater across cities, education groups, or age groups; whether the main effects of city, education, and age on log income are independent of one another; how much inequality can be accounted for by each of the explanatory variables; and how important are variations across these groupings as compared with the variations within them. The explanatory variables are:

City: Bogota, Barranquilla, Cali, Medellin

Education of head of the household: None, primary (some or all), secondary (some or all), higher (some or all)

Age of head of the household: Less than 35, 35-49, 50-64, 65 and over.

---

[1] For a similar analysis for all of Colombia, see Fields and Schultz (1977). The computer software used is the ANOVA program in the Statistical Package for the Social Sciences (SPSS). The SPSS manual contains a clear description of analysis of variance procedures by Kim and Kohout (1975) to which readers unfamiliar with the technique are referred.

[2] See, respectively, Fields (1976), Fields and Marulanda (1976), and Heady (1976). Both market and institutional reasons for earnings differentials are considered in these studies.

Table 6 presents the results of the inequality decomposition by income-determining factors. Looking first at the main effects, each explanatory factor helps account for inequality. The significance column shows that each of these effects is statistically significant at the .001 level. However, the contributions of the three sets of factors are by no means equal. Of the 36.9% of the log variance explained by the main effects, education accounts for nearly all of it, 34.7%. By contrast, age accounts for just 4.2% and city 0.4%. Education thus overwhelms the other explanatory factors. One way of interpreting these results is this: if you wanted to ask one question of a family to ascertain its economic position, you would be much better able to predict income if you asked about the education of the family head rather than the age or city of residence.

Immediately below the main effects in Table 6 are the interaction effects. The education-city interactions, for example, allow for the possibility that the effect of education on income might depend on which city one lives in or alternatively that the effect of city on income might depend on one's level of education. The three sets of two-way interaction effects --- city-education, city-age, and education-age --- together add significantly to the explanation of inequality, but they account for only 1.6% of the log variance. Thus, the explanatory effects of education, age, and city are not independent of one another, but the degree of interdependence is small. Whether the 1.6% additional explanatory power contributed by the two-way interaction warrants a quadrupling of the number of explanatory categories from 9 to 36 is a matter of some economic judgment. The three-way interactions, however, contribute even less explanatory power, only 0.5%. Even on narrow statistical grounds, their inclusion is not justified.

Another useful output of the ANOVA program used is a multiple classification analysis (MCA). The MCA exploits the formal equivalence between the linear model used in analysis of variance and the linear model used in multiple regression analysis, producing estimates of the quantitative effect of each category of each explanatory factor, expressed as deviations from the grand mean of the logarithm of income (6.52). These estimates appear in the second block of Table 6. The first column gives the gross effects of membership in a particular category, unadjusted for any other explanatory variable. For example, persons with no education on average earn 74% less than the overall mean and persons with higher education 90% more. The second column gives marginal effects which do adjust for the influence of other variables. The corresponding marginal effects are 82% less than the overall mean for the uneducated and 93% more than the overall mean for the highly-educated. The adjusted effects are greater in absolute value than the unadjusted ones. This means that education is negatively related to some other explanatory factor. That factor is age. In Colombia, as elsewhere, young family heads tend to be better-educated. The unadjusted comparisons do not allow for this fact. Since the better-educated group includes disproportionately many young workers at the early stages of their careers, the unadjusted comparisons understate the income gain that a representative individual would realize if he or she had more education. Likewise, the adjusted age effects are greater absolutely than the unadjusted ones, these steeper age-income profiles arising for the same reason: the unadjusted comparisons take no account of the disproportionately large number of young persons who are relatively well-educated and who consequently move along

Table 6

Decomposition of Inequality in Urban Colombia by Income Determinants, 1967-68

Decomposition of Log Variance

Source of Variation	Sum of Squares	F (df)	Significance of F
Main Effect Explained:			
City	9.2 (0.4%)	5.74 (3)	.001
Education	876.4 (34.7%)	546.1 (3)	.001
Age	106.3 (4.2%)	66.2 (3)	.001
Covariance	- 58.9 (-2.3%)		
Total, Main Effects	933.0 (36.9%)	193.8 (9)	.001
Two-Way Interactions Explained:			
City-Education	13.3 (0.5%)	2.76 (9)	.003
City-Age	4.3 (0.2%)	.90 (9)	.999
Education-Age	21.9 (0.9%)	4.54 (9)	.001
Covariance	1.4 (0.0%)		
Total, Two Way Interactions	40.9 (1.6%)	2.83 (27)	.001
Three-Way Interactions Explained:			
City-Education-Age	13.0 (0.5%)	.90 (27)	.999
Total Explained	987.0 (39.0%)	29.3 (63)	.001
Unexplained	1542.3 (61.0%)		
Total	2529.2 (100.0%)		

Multiple Classification Analysis

Grand Mean = 6.52	Unadjusted Effects	Adjusted Effects
City Effects:		
Bogota	.09	.09
Barranquilla	-.02	-.06
Cali	-.05	-.02
Medellin	-.03	-.01
Education Effects:		
None	-.74	-.82
Primary	-.43	-.44
Secondary	.36	.37
Higher	.90	.93
Age Effects:		
Less than 35	-.18	-.27
35-49	.05	.06
50-64	.08	.16
65 and over	.25	.38

Proportion of Log Variance Explained

$R^2 = .390$

different income paths than the less-educated. Besides revealing these covariations, the MCA coefficients are of considerable interest in and of themselves in quantifying the differentials associated with various income-determining factors.

Overall, the main effects and interaction effects together account for 39.0% of the variance in the logarithms of income. This means that 39.0% of inequality is attributable to income variation across education-age-city groups, the remainder due to variation within these groups. As compared with research on other countries (e.g., that of Mincer (1974) on the U.S.), this is a very good start toward explaining inequality. Psacharopoulos (1973), Blaug (1973) and others have emphasized education's role in explaining income and income inequality in less developed countries. In the case of Colombia, this concentration seems fully warranted.

Part of the remaining variation within groups is due to the use of education and age categories rather than years. In Colombia, each year of primary education increases income on average by 20%. Persons who complete primary education (5 years) therefore receive about twice the income of persons who complete just one year. By merging these individuals with different years of education into a single category of "primary educated," some information loss occurs. A quantitative estimate is found in the work of Fields and Schultz (1977), who find that in Colombia the proportion of variance explained by continuous education and age data rather than discrete groupings is about 10% higher.

Some other part of the within-group variation is due to the limited number of income determinants considered. Among the other factors known to explain family incomes in Colombia are: the number of workers in the family and their educational, age, and sex distribution; migration histories;

employers' characteristics; parents' socio-economic position; etc. In future research, allowance for the effects of these factors would undoubtedly increase the percentage of inequality accounted for.

Finally, some part of the within-group variation is due to simple luck. We cannot possibly hope to account for all income variability in a stochastic world. It will be interesting to see how far future researchers will be able to go toward accounting for Colombian inequality.

VI. Conclusions

This paper has examined income inequality in urban Colombia, decomposing overall inequality according to functional, geographical, and income-determining factors. The statistical results provide a factual basis in an area of critical importance to the study of economic development, one in which only a handful of rigorous empirical research studies are to be found.

In respect to a functional accounting for overall inequality, the Colombian data, in common with recent and as-yet unpublished analyses of Taiwan and Pakistan, reveal the prime importance of labor income. Labor income accounts for almost 70% of total inequality in urban Colombia. Very simply, most people get most or all of their incomes from the work they do. True, other income sources, particularly capital, are more unequally distributed. Yet, precisely because of their high concentration and because of their small functional shares, these other sources account for less overall inequality than does labor income. If only ten or twenty percent of the people receive any appreciable amount of income from wealth, income inequality among the remaining eighty or ninety percent must be explained otherwise. That explanation has something to do with the fifty to one ratio of earnings between doctors, lawyers, and other professionals on the one hand and the domestic workers whom they employ on the other.

Unlike other research studies in this area, which have made use of aggregated tabulations of total incomes and incomes from the various

functional sources, the Colombian research is based on micro data on individual families. We observed the results of an experiment in which the micro data were aggregated as in the tabulations for other countries and all decomposition statistics were recomputed. The overall Gini coefficient of inequality, the factor income shares, and the factor inequality weights exhibit only minor differences. Thus, the conclusions reached in past studies of other countries regarding the importance of labor income in accounting for overall inequality are sustained. Where the use of aggregate data distorts the true patterns is in decomposing the factor inequality weights. The true correlations between factor incomes and total incomes are overstated when aggregate data are used and the true factor Gini coefficients understated, the degrees of overstatement or understatement ranging from 35% to 280%. Previous researchers, who had access only to aggregate data, could not have known the serious magnitudes of the biases which arise in the type of aggregated data employed. However, future researchers wishing to decompose inequality along these lines would be well-advised to work with micro data.

Turning to other types of inequality decompositions, regional inequality is often suspected as a major contributor and is so blamed in Colombia. Although average incomes differ across the sample cities by some 30%, less than 1% of overall inequality is found to be associated with income variation <u>across</u> cities. 99+% of inequality in urban Colombia is due to variations <u>within</u> cities. An explanation for the within-city variation must be sought.

A large part of the answer lies in labor force heterogeneity. Workers differ by education and age and receive correspondingly different rewards. Nearly 40% of inequality in Colombia is found to be explainable in terms of differences by education, age, and city. Almost all of this explained component is attributable to educational differences (35%). Age contributes only a small amount (4%) and city even less (<1%)

At a deeper level, it might be asked: Why does each explanatory factor account for what it does? Take education, for example. Why do persons with higher education earn so much more than illiterates? Is the return to education a return to human capital acquired through schooling or does it result from meritocratic admission procedures in the schools, the buying of scarce spaces by rich parents, the payment of higher salaries to well-educated employees out of proportion to productivity differentials, or some other cause? We are disturbingly far from understanding the basic determinants of incomes and the root causes of income inequality, in Colombia or elsewhere.

BIBLIOGRAPHY

Ayub, M., Income Inequality in a Growth-Theoretic Context: The Case of Pakistan, unpublished doctoral dissertation, Yale University, 1977.

Berry, R.A. and Urrutia, M., Income Distribution in Colombia, Yale University Press, New Haven, 1976.

Blaug, M., Education and the Employment Problem in Developing Countries, Geneva, International Labour Organisation, 1973.

Fei, J.C.H. and Ranis, G., "Income Inequality by Additive Factor Components," Economic Growth Center, Yale University, Center Discussion Paper No. 207, June, 1974

Fei, J.C.H., Ranis, G., and Kuo, S.W. Equity with Growth: The Taiwan Case, unpublished manuscript, 1977.

Fields, G.S., "Decomposing LDC Inequality," Economic Growth Center, Yale University, Center Discussion Paper No. 263, June 1977.

Fields, G.S., "Education and Economic Mobility in a Less Developed Country," Economic Growth Center, Yale University, Center Discussion Paper No. 237, revised version, June, 1976.

Fields, G.S. and Jaramillo, H., "A Guide to the Use of Microeconomic Data Sets in Colombia," Economic Growth Center, Yale University, September, 1975.

Fields, G.S. and Marulanda, N., "Intersectoral Wage Structure in Colombia," Economic Growth Center, Yale University, Center Discussion Paper No. 251, August, 1976.

Fields, G.S. and Schultz, T.P., "Sources of Income Variation in Colombia: Personal and Regional Effects," Economic Growth Center, Yale University, Center Discussion Paper No. 262, June, 1977.

Heady, Christopher, "Alternative Theories of Wages in Less Developed Countries: An Empirical Test," Economic Growth Center, Yale University, Discussion Paper No. 254, October, 1976.

Isaza, R. and Ortega, F., "Eucuestas Urbanas de Empleo y Desempleo: Análisis y Resultados," CEDE, Monograph No. 29, 1971.

Kim, J. and Kohout, F.Y., "Analysis of Variance and Covariance: Subprograms Anova and Oneway," Chapter 22 in Norman H. Nie, et. al., Statistical Package for the Social Sciences, 1975.

Kuznets, S., "Quantitative Aspects of the Economic Growth of Nations: VIII. Distribution of Income by Size," Economic Development and Cultural Change, Part II, Vol. 11, No. 2, January, 1963.

Mincer, J., Schooling, Experience and Earnings, New York, Columbia University Press, 1974.

Musgrove, P., "Urban Household Income and Consumption Patterns in Latin America: A Comparative Analysis of Colombia, Paraguay, Peru, and Venezuela," Washington, Brookings Institution, mimeo, June, 1974.

Prieto, R., Estructura del Gasto y Distribución del Ingreso Familiar en Cuatro Ciudades Colombianas: 1967-68, Centro de Estudios sobre Desarrollo Económico, Universidad de Los Andes, Mayo, 1971.

Psacharopoulos, G., Returns to Education: An International Comparison, San Francisco, Jossey Bass/Elsevier International Series, 1973.

Williamson, J., "Regional Inequality and the Process of National Development: A Description of the Patterns," Economic Development and Cultural Change, XIII, 4, part II, July 1965.