MALE-FEMALE INCOME DISTRIBUTIONS IN FOUR CANADIAN METROPOLITAN AREAS: AN APPLICATION USING PERSONAL INCOME TAX DATA

C.Dagum, G.Grenier, University of Ottawa; D.Norris, M.Bédard, Statistics Canada

1. INTRODUCTION

Traditionally, analyses of the evolution over time of income distribution and inequality have been limited to the study of large geographical areas, such as countries or major regions of countries. The main reason for this limitation was that these analyses were conducted with data obtained from surveys carried out annually with relatively small sample sizes. Analyses of small areas were possible using census data, but these could be done only once every ten years.

Recently, data from income tax records have been made available for small geographical areas on an annual basis by Statistics Canada. This new source of information opens a wide range of possibilities for economic research on incomes. In this paper, we explore one of these possibilities by analysing male-female income distributions and income differentials over the period 1969-1981 in four selected Canadian metropolitan areas which experienced different economic trends over that period. A parametric model of income distribution is fitted to employment income data by sex and age group, and measures of inequality within and between distributions are estimated. The impact of some socio-economic variables upon these inequality measures is also investigated. The results indicate that the model of income distribution and the data base which are used provide useful tools for the analysis and economic interpretation of income inequalities.

2. THE DATA BASE

The source files for this study are derived from the T1 individual income tax records of Revenue Canada Taxation. For the years 1977-1981, the data were obtained from the 100 percent tax files; for earlier years, the data were derived from a 10% sample of all taxfilers. For 1977, data were produced on both a sample basis and a 100% basis and the results were compared to assess the validity of the sample data. The results indicated that there were small differences between the two files. However the differences were in general not significant and the two sources taken together provide consistent observations of the male and female employment incomes and consistent time series of statistics such as the Gini ratio, the male-female income differential ratio, and the estimated mean and median. As expected, because of sample fluctuations, the sum of squared deviations between the fitted and the observed data from the 10% sample were larger than that between the fitted and the observed data from the 100% files.

The personal income tax file has a high coverage of the adult population, especially the labour force. For tax year 1981, approximately 80.2% of the population 15 and over filed a tax return. An earlier study (Norris and Mussely, 1983) compared the number of persons reporting employment income, a proxy for labour force participation, as reported in the tax records to corresponding census data. The results showed that for the population 25 and over the two sources yielded near identical counts although the tax records were approximately 10% lower for persons under age 25.

Other strengths of the tax records include the availability of basic demographic information (age, sex and marital status) and detailed information on income by sources. In addition, the tax records contain detailed mailing addresses that allow one to produce data for small geographic areas.

An important shortcoming of the tax file is that individuals with low incomes may not be required to file a tax return. The comparison to census data for 1980 showed that there were approximately 10% fewer persons with employment income less than $3000 in the tax records compared to the census; for incomes in excess of $3000, the difference was only 1%. Finally, because of frequent changes in tax laws, there may be problems with the inter-temporal comparability of statistics derived from tax records. An important change occurred in 1978 with the introduction of the Child Tax Credit Program. In the case of married couples, this program pays a child tax credit generally to the mother, but to apply for the credit, it is necessary to file a tax return (stating family income) even though the potential recipient had no sources of income. The introduction of this program increased the number of filers by 10% and while many of these new filers reported no income, a number of them reported small amounts of employment income. This contributed to an increase of the low employment income frequency and with this an associated increase in the intra-income inequality.

3. SELECTION OF STUDY AREAS

This research is part of a larger study that is directed at the development and evaluation of income data produced from tax records. As such, one of the goals of the study was to utilize income distribution theory and models to investigate the potential of using tax data for the socio-economic analysis of small areas. Although the Census of Population provides detailed income data, this is only available every ten years. Annual income surveys are also carried out, but they contain very limited information at the sub-provincial level. Therefore, in choosing the study areas it was decided to focus on smaller census metropolitan areas. Four such metropolitan areas were identified: Chicoutimi-Jonquière (Québec), London (Ontario), Saskatoon (Saskatchewan) and Sudbury (Ontario). These were chosen to provide a range of areas that differed in their socio-economic profile and that have experienced different economic and demographic trends during the decade of the 1970's, the core study period.

Some general background information about the four metropolitan areas is given in Table 1. This information is presented for both the census years 1971 and 1981, and, as a basis for comparison, the same information is also presented for Canada as a whole.

First, we note that the four areas, whose population in 1981 ranged from 137.2 thousands in

Chicoutimi-Jonquière to 283.7 thousands in London, experienced different population growths during the seventies. The area with the highest population growth is Saskatoon, with a rate of 22%, almost twice the rate for Canada. In contrast, Sudbury saw its population decrease by 4.9% during the same period, while those of Chicoutimi-Jonquière and London grew at moderate rates [1].

The regions are also different from each other with respect to the age structure of their populations. In both census years, Chicoutimi-Jonquière and Sudbury had a relatively young population, compared to the other areas and to Canada as a whole. We note that the Canadian population became substantially older during the seventies and that is also true for our four metropolitan areas. The increase in the proportion of the population which is aged 65 years or more was particularly significant in Sudbury, where we noticed that the total population growth was negative. In constrast, in Saskatoon, whose total population grew quite rapidly, the proportion of the population aged 65 years and more increased only slightly.

Variations among the areas are also found in the labour force participation rates. In particular, we notice that Chicoutimi-Jonquière has lower rates than other regions for both males and females in both census years. Between the two census years, female participation rate increased in all the areas. For men, the patterns differ across areas: participation rates increased between 1971 and 1981 in Chicoutimi-Jonquière and Saskatoon, while it decreased in Sudbury and remained approximately the same in London.

The last characteristic of the four regions that is considered in Table 1 is unemployment. Again, Chicoutimi-Jonquière differs from the other areas by its quite large unemployment rate for both sexes. We also notice a particularity in Sudbury whose unemployment rate for women is very large relative to that of men. In the other areas, the unemploymewnt rate for women is only slightly higher than for men. For Canada as a whole, the unemployment rate was a little bit lower in 1981 than in 1971. We note, however, that the areas present different directions of changes in the unemployment rate. The unewployment rate increased significantly in Sudbury and decreased significantly in Saskatoon. Interestingly, these are the two regions which experienced the two extreme population patterns: Sudbury declining and Saskatoon growing.

Finally, it may be important to note that two of the four metropolitan areas under study are each dominated by one industry. In Sudbury it is the copper-nickel mining sector which dominates the economy of the region, while in Chicoutimi-Jonquière, this role is played by the aluminum industry. The region of Saskatoon is also influenced by the potash mining industry.

4. THE INCOME DISTRIBUTION MODEL AND THE INTER- AND INTRA- INCOME DISTRIBUTION MEASURES OF INEQUALITY

The income distribution model specified by Dagum {1},{2},{6} is fitted to the observed income distributions. The non-linear least square method of parameter estimation is applied, using an iterative algorithm that searches the minimization of the sum of the squared deviations of the actual from the fitted values {5}. The mathematical form for Type I model is:

$$(1)\quad F(x)=(1+\lambda x^{-\delta})^{-\beta}, \qquad (\beta,\lambda,\delta)>0,\ x>0,$$

and for Type II, $(0<\alpha<1)$, and Type III $(\alpha<0)$ models,

$$(2)\quad F(x)=\alpha+(1-\alpha)(1+\lambda x^{-\delta})^{-\beta}, \quad (\beta,\lambda,\delta)>0,$$

such that $x>0$, if $0<\alpha<1$; $x>x_0>0$, if $\alpha<0$, where x satisfies the equation $F(x_0)=0$.

The Gini ratio is obtained from the fitted income distributions. It is a measure of the employment income inequality **within** each population, i.e., an intra-income distribution measure of inequality.

To measure the male-female income differential ratio, i.e., an inter-income distribution measure of inequality, for each region, by age groups, and the employment income differential between regions, by sex, from 1969 to 1981, an economic distance ratio D introduced by Dagum {3}, {4} is used. That is,

$$(3)\quad D=(d-d^*)/(\Delta-d^*),$$

which is a normalized and dimensionless measure of inequality between distributions.

Before giving the mathematical definition of the symbols specified on the right-hand-side of (3), let us introduce the notations to be used.

Given two populations of income recipients, the one with higher mean income is defined as the more affluent. To this, the subscript 2 is attached whereas the less affluent population bears the subscript 1. Hence, for $i=1,2$, X_j stands for the income variable of the i-th population; $F_i = F_i(x) = P(X_i \leq x_i)$ stands for the cumulative income distribution (CDF), and $E_i(X) = E(X_i)$ for the mathematical expectation or mean income of the i-th population. Hence, when the subscript i, i=1,2, is attached to the mathematical expectation operator, it means that the mathematical expectation is taken with respect to (i.e., the weighting factor is) the i-th distribution function.

Bearing in mind the above conventions, we define the symbols on the right-hand side of (3).

$$(4)\quad d = E(X_2-X_1 \mid X_2>X_1,\ E(X_2)>E(X_1))$$

$$= E_1(XF_2) + E_2(XF_1) - E_1(X),$$

$$(5)\quad d_i^* = 2E_i(XF_i) - E_i(X),\ i=1,2,$$

i.e., it is the value of d under the null hypothesis of identically distributed income variables, which is represented by the CDF $F(x)$, and

$$(6)\quad d^* = \min(d_1^*, d_2^*).$$

The symbol Δ stands for the Gini mean difference between two distributions, i.e.,

$$(7)\quad \Delta = E(|X_1 - X_2|).$$

It follows from (3)-(7) and the assumption $d_1^* < d_2^*$ that

$$(8) \qquad D = \frac{E(XF) + E(XF) - 2E(XF)}{2E(XF) + 2E(XF) - 2E(XF) - E(X)}$$

This income differential or economic distance ratio D can be obtained from the fitted income distributions (parametric estimate), or from the observed income distributions (distribution-free estimate). For the models specified in (1) and (2), the formulas corresponding to the mathematical expectations on the right-hand side of (8) are presented in Dagum {3}.

5. MALE-FEMALE INCOME DIFFERENTIAL, AND MALE AND FEMALE INCOME INEQUALITY BY REGION

The Dagum model of income distribution was fitted to the male and female employment income distributions by size for each of the four regions under inquiry and for the following age groups: (i) males and females with 20 to 29 years of ages; (ii) males and females with 30 to 54 years of ages; (iii) males and females 15 years and over (all ages).

The other age groups (under 20 and 55 and over) were not considered separately because of the small sample sizes, which would not allow a reliable estimate and analysis. The model was fitted for the odd numbered years only between 1969 and 1981, with the exception of 1970 and 1980, in view of possible future comparisons with census data.

The income differential ratio D was calculated for the male-female income distributions for each age group by region (Figures 1-4). The Gini ratio of each distribution by region and sex was also calculated (Figures 5-8).

The male-female income differential or Dagum ratio for the three age groups under study follows a mild downward time path for London, and a stationary time path for the other three regions, with the only exception of the 20-29 age group in Chicoutimi-Jonquière, which presents a mild upward time path. A common trait for the four regions under study is that the 20-29 age group presents the smallest male-female Dagum ratio during the entire period. It is followed by the all ages male-female ratio and the highest level corresponds to the 30-54 age group, with the exception of Chicoutimi-Jonquière in 1970, where the 30-54 age group male-female ratio was smaller than that for all ages. However, the male-female ratios for all regions, age groups and years are extremely high, underlying a large income disparity between males and females. One important reason is the labour market segmentation by occupation, whereby a relatively large proportion of the female labour force is employed in relatively low paying occupations, such as clerical, sales and service. These occupations are also characterized by an important participation of part time employees. A more basic reason for this labour market segmentation can be found in the human capital embodiment of the male and female labour forces. Finally, the prime age of the adult labour force (30-54 age group) systematically presents the highest male-female income differential. As expected, the null hypothesis of no income differential between males and females, i.e., H : D=0, is rejected in every case study, even at the 0.0001 significance level.

It follows from Figures 5-8 that the female intra-income inequality, as measured by the Gini ratio, is systematically higher than the male intra-income inequality for each age group, and in general, they cluster above the Gini ratio corresponding to the male income distributions. With the exception of Saskatoon which presents a stationary time path for the male and female Gini ratios by age group, the other three regions show a mild upward trend, i.e., a worsening of the income inequality by sex and age group. Although there is some crossing of the Gini ratio time paths, the highest in every region corresponds to the total female population, followed by females in the 30-54 age group and by females in the 20-29 age group.

The time path of the male Gini ratio by region shows a different pattern. In Chicoutimi-Jonquière, the Gini ratio time path of the total male population is completely bounded between the Gini ratio time paths of the 30-54 and the 20-29 age groups. In London and Saskatoon, the time path of the Gini ratio corresponding to the total population of males is systematically above the male Gini ratio for the 20-29 age group which in turn is above male Gini ratio for the 30-54 age group. Sudbury follows similar trends with the exception of the 1979

Figure 1. Dagum Ratios by Age Groups: Chicoutimi-Jonquière

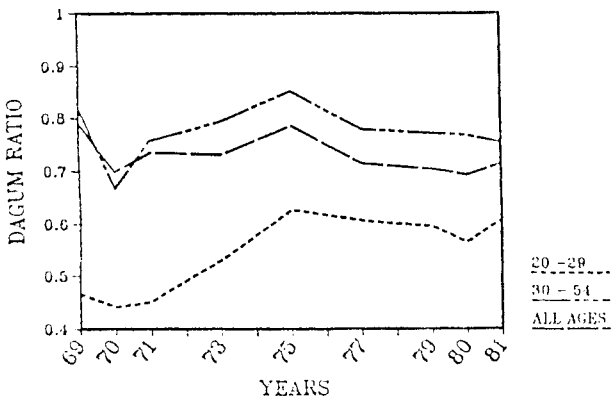


Figure 2. Dagum Ratios by Age Groups: London

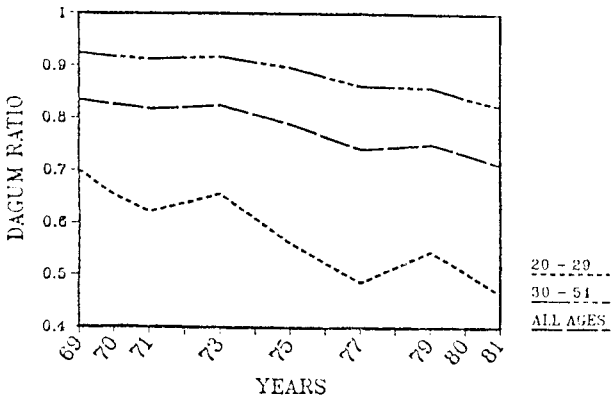


Figure 3. Dagum Ratios by Age Groups: Saskatoon

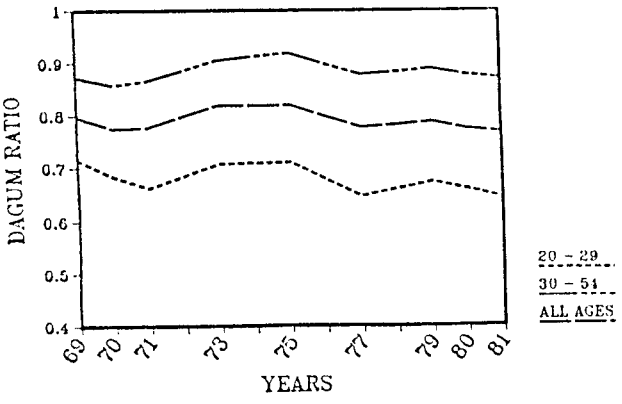


Figure 4. Dagum Ratios by Age Groups: Sudbury

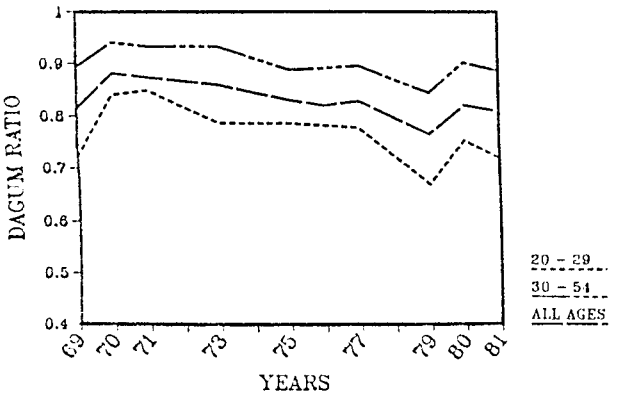


Figure 5. Gini Ratios by Sex and Age Groups: Chicoutimi-Jonquière

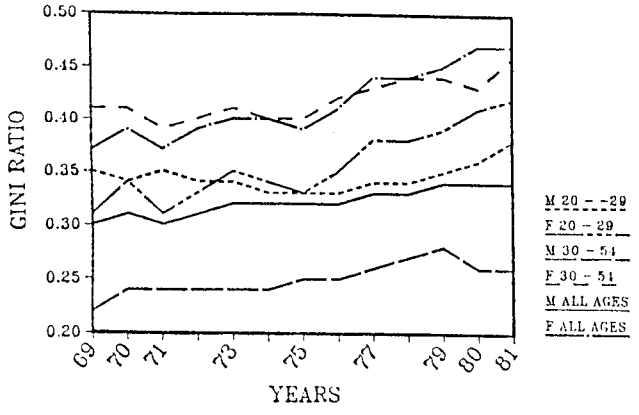


Figure 6. Gini Ratios by Sex and Age Groups: London

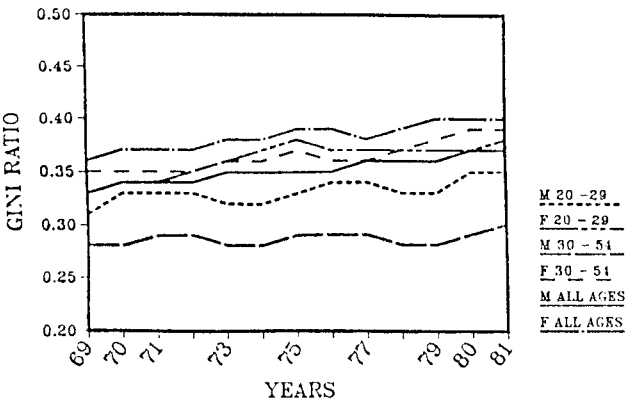


Figure 7. Gini Ratios by Sex and Age Groups: Saskatoon

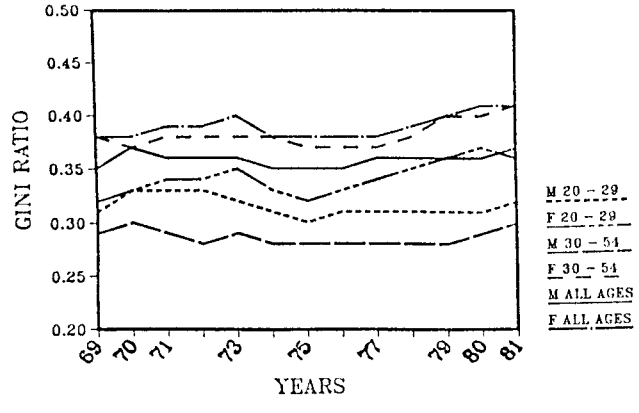
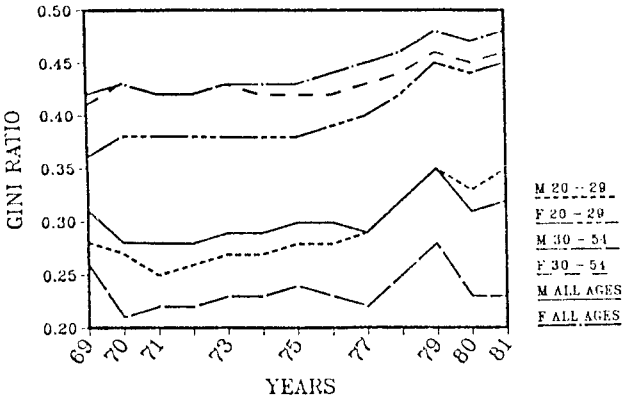


Figure 8. Gini Ratios by Sex and Age Groups: Sudbury



and 1981 years, where the male Gini ratio for the 20-29 age group is above the total male population.

Unlike the female Gini ratio, the male Gini ratio for the 30-54 age group has a time path below all the other and presents a relatively low level of income inequality. This characterizes a stable adult labour force, and a stable occupational structure and human capital distribution, whereas the female labour force is in a dynamic process of shaping new occupational and income structures.

6. TOWARDS AN ECONOMIC EXPLANATION OF THE INCOME INEQUALITY TIME PATHS

The evolution over time of the income differential and inequality measures in each of the four metropolitan areas can be affected by changes in

economic conditions which are related to trends and cycles. In this section, we consider this question by analysing relationships between these measures and some socio-economic indicators.

Given the relatively short length of the time period under consideration (12 years), and the small number of observations (only 9 data points), this analysis has to be of an exploratory nature since we obviously cannot estimate a complete model of the determinant of the evolution of income differentials and inequalities. Our investigation is also seriously limited by the availability of annual data at the metropolitan area level. Given these constraints, we decided to do the analysis on the basis of simple correlations between the Dagum and Gini ratios on the one hand and the socio-economic indicators on the other. The evolution of economic conditions over time was evaluated from several proxies. Because of the scarcity of data at the regional level, some of these variables are at the national level. The other variables that we consider refer to economic conditions within each region, or to industries which include an important part of their activities in these regions.

At the national level, the variables considered are inflation, GNP and some of its components such as personal expenditures, gross fixed capital formation and government expenditures, unemployment rates by sex and for two age groups. At the regional level, we used an index of employment in larger firms [2], an indicator of weekly earnings, the number of building permits, several indicators of the value of construction and some indicators of employment in the manufacturing industry. For the Chicoutimi-Jonquière region, we used, as another indicator, employment in the smelting and refining industry in Québec, as a proxy for activity in the aluminum sector. Similarly, for the Saskatoon region, we used an indicator of employment in the potash mining in Saskatchewan. Finally, for the Sudbury region, we used an indicator of production of nickel-copper ore in Canada.

After obtaining data for these variables for each of the years under consideration, we computed correlation coefficients between each indicator and the Dagum and Gini ratios, using both the levels and first differences of the variables. Correlations on first differences were used in order to purge the variables from the trends. At present, only the Dagum and Gini ratios for total male and female age groups were considered in the analysis [3].

In general, as we expected to some extent, we did not find extremely strong correlations which would suggest very clear-cut causality links between these indicators and income differential and inequality measures. For instance, many of the correlations that we observed simply reflected general trends which do not imply that the variables are related. In a few cases, however, the observed correlations may have some economic interpretations. The discussion below summarizes the major results of the analysis.

For the Chicoutimi-Jonquière region, we found positive correlations between the male and female Gini ratios and variables measured at the national level, such as prices, GNP and unemployment. But these appear to be due to the upward trend in all these variables and in the Gini ratios, since these correlations disappear when variables are taken in the first differences. The Dagum ratios between male and female income distributions do not show any trend either upward or downward and, as a result, there is no significant correlation between it and any of the national variables. Among the variables measured at the regional levels, positive correlations were found between the male and female Gini ratios and employment, value of construction and activity in the smelting and refining industry (aluminum), but again these are due mainly to trends. When the same variables are taking in the first differences, there is very little correlation. The only result which may be of some interest is a positive correlation between male manufacturing employment and the Dagum ratio between males and females.

For the London region, we also found positive correlations between the male and female Gini ratios, and the variables measured at the national level, which are due to a general upward trend in these variables. On the other hand, since the Dagum ratio shows a downward trend, it is negatively correlated with the same variables. When the variables are taken in the first differences, the only correlation that remains is a negative correlation between the unemployment rate and the income differential between males and females. Among the regional variables, we found some correlations between the income inequality measures and earnings and value of construction, but these again seem to be mainly trends. With variables taken in the first differences, we found, as we did in the Chicoutimi-Jonquière region a positive correlation between manufacturing employment and the income differential between males and females. This may be explained by the fact that manufacturing is dominated by male employment and that increases in the levels of activity of that sector affect positively male income relative to female incomes, and vice-versa.

Very few high correlations were found for the Saskatoon region. The major reason is that the Gini and Dagum ratios do not show as clear trends as in the above cases. It may be interesting to note that, with variables in the first differences, a negative correlation was found, as in London, between the national unemployment rate and the male-female income differential ratio. There is also a negative correlation between the manufacturing employment and the male Gini ratio.

Finally, for the Sudbury region, only the female Gini ratio is positively correlated with the national variables, which is due again to a general positive upward trend in the female Gini ratio, while there is no general trend in the Dagum ratio and the male Gini ratio. As far as regional factors are concerned, the evolution of the income inequality measures can be related to the evolution of the copper-nickel mining industry. Besides being characterized by a general decline in activity during the decade, this industry also experienced some fluctuations. For instance, production hit a low in 1979, due to a strike, before increasing slightly during the

following years. It is interesting to note that the Dagum and Gini ratios also showed a particular behaviour in 1979, the male-female income differential ratio decreasing slightly, and the Gini ratios increasing, especially for males. Partly as a result of this, we found significant correlations between the inequality measures and copper-nickel production and also the employment variables, in both levels and first differences. More precisely, the Dagum ratio is positively correlated with these variables and the male Gini ratio is negatively correlated. This may be explained by the fact that the copper-nickel mining industry employs mainly males, and when things got bad in that industry (workers were laid-off or went on strike), male incomes decreased relative to female incomes, thus decreasing the Dagum ratio. Also, since the annual incomes of those who keep their job do not usually fall and are in general higher than the incomes of those who lost theirs (less skilled workers are laid-off first), inequality within the male group then increased.

To conclude, our analysis suggests that there are some interesting relationships between income differentials and inequality measures and socio-economic indicators. However, we have to be careful in the interpretation of these results. In our research in progress, we plan to investigate these relationships further.

TABLE 1. SOME CHARACTERISTICS OF THE SELECTED METROPOLITAN AREAS (CMA)

	Chicoutimi-Jonquière	London	Saskatoon	Sudbury	Canada
Population 1971 ('000)	126.4	253.0	126.5	157.7	21,568.3
Population 1981 ('000)	137.2	283.7	154.2	149.9	24,343.2
Change 1971-1981 (%)	8.5	12.1	22.0	-4.9	12.9
Age structure (%)					
1971: 0-14	32.5	27.7	29.1	33.1	29.6
15-64	63.3	63.6	61.9	62.9	62.3
65+	4.2	8.7	9.0	4.0	8.1
1981: 0-14	24.0	20.9	22.2	24.7	22.5
15-64	70.0	69.0	68.3	68.2	67.8
65+	6.0	10.1	9.5	7.1	9.7
Labour Force Participation Rate (%)					
1971: Males	66.6	81.0	78.7	82.5	76.4
Females	28.1	47.3	44.4	37.3	39.9
Total	47.2	63.5	60.8	60.9	58.0
1981: Males	74.3	81.0	81.5	75.8	78.2
Females	39.3	58.3	57.9	47.9	51.8
Total	56.6	69.1	69.1	61.7	64.8
Unemployment Rate (%)					
1971: Males	14.7	5.7	8.1	4.5	7.3
Females	15.5	7.5	9.7	10.0	8.8
Total	15.0	6.5	8.7	6.2	7.9
1981: Males	13.5	5.7	5.3	6.4	6.5
Females	20.0	7.1	7.0	10.8	8.7
Total	15.8	6.3	6.1	8.2	7.4

Source: Census of Canada, 1971 and 1981.

REFERENCES

{1} Dagum, C., "A New Model of Personal Income Distribution: Specification and Estimation", **Économie Appliquée**, XXX(3), 1977, 413-436.

{2} Dagum, C., "The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio", **Économie Appliquée** XXXIII (2), 1980, 327-367.

{3} Dagum, C., "Inequality Measures Between Income Distributions With Applications", **Econometrica**, 48 (7), 1980, 1791-1803.

{4} Dagum, C., "Male-Female Income Differential in Canada", **Proceedings of the American Statistical Association, Business and Economics Section**, 142nd Meeting, 1982, 426-430.

{5} Dagum, C., Chiu, K., Gunaratne, L.L. and Barletti, B., **Econometric Package for Income Distribution (EPID)**, University of Ottawa and Time Series Research and Analysis Division, Statistics Canada, 1983.

{6} Dagum, C., "Income Distribution Models", **Encyclopedia of Statistical Sciences**, Vol.IV. Johnson, N.L. and Kotz (eds.). New York: John Wiley and Sons, 1983, 27-34.

{7} Norris, D. and Mussely, A., "A Measure of Annual Work Experience Using Individual Income Tax Records — Some Initial Results". In **Statistics of Income and Related Administrative Record Research: 1983.** Department of the Treasury. Internal Revenue Service. Statistics of Income Division, October 1983, 215-219.

FOOTNOTES

[1] These growth rates are based on definitions of census metropolitan areas which are standardized for both census years.

[2] Since 1976, the unemployment rate is published at the metropolitan area level.

[3] The tables showing the results of the analysis are available from the authors on request.