

Decomposition of Income Inequality

Using an Earning Equation

By

Stéphane Jetté

616158

Major Paper Presented to

The Department of Economics of the University of Ottawa

In Partial Fulfillment of the Requirements of the M.A. Degree

Supervisor: Professor Gilles Grenier

ECO 7997

Ottawa, Ontario

April 2002

Preface..... iv

Acknowledgements..... v

Abstract..... vi

Introduction..... 1

1 Inequality 2

 1.1 Inequality Measurement Issue 3

 1.2 Previous studies on inequality sources 4

2 Methodology – Decomposition by factor and time 7

 2.1 Decomposition by explanatory factor..... 8

 2.2 Decomposition by attribute and structural effect over time..... 10

3 Decomposition Application – Data, Earning Equation and Results 12

 3.1 The data..... 12

 3.2 The regression model for the earning equation..... 16

 3.3 Gini coefficient inequality decomposition by factor 18

 3.4 Gini coefficient inequality decomposition by structure and attribute over time 22

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

3.5 Results sensitivity	25
4 Findings and Policy Implications.....	25
4.1 Education	26
4.2 Gender.....	28
4.3 Presence of children.....	29
Conclusion	31
Bibliography	32
Appendix I Review of Inequality Measures	35
Appendix II Earning Equation Model Definition	37

Preface

In January 2001, the IRPP/CSLS¹ organised a conference titled "Linkages between Economic Growth & Inequality". The conference had two objectives: first, to present the latest research findings on that topic; and second, to debate policies affecting growth and inequality, based upon conclusions drawn from the research. While the organisers clearly expressed their view that reductions in inequality positively influence economic growth, there were a wide variety of presentations, with varying and sometimes conflicting conclusions.

By the end of the conference, one could infer that inequality could be beneficial when viewed as a reward for personal achievements due to merit. This type of inequality may stimulate the drive for people to improve their skills for greater personal and monetary gain. On the other hand, inequality of opportunity is viewed as a negative consequence of unequal social and economic status within society. Conventional wisdom assumes that in the long term, economic growth leads to a more equitable society. Altman (2001) concludes that convergence toward equity, however, has not been observed. The question then becomes, what are the underlying factors of inequality and how they can be influenced by governmental policy?

¹ Institute for Research on Public Policy and Centre for the Study of Living Standards

Acknowledgements

A first version of this paper was produced as research project done under Ottawa University's course ECO-6904 SELECTED TOPICS IN APPLIED ECONOMICS. This paper reviews alternative methodologies and pushes further the analysis in order to deepen the theoretical background and the robustness of the conclusions. The opinions and conclusions are solely those of the author and should not be construed as representing the opinions or policies proposed by the Department of Finance Canada, the Canadian Government, or the University of Ottawa. The policy implications presented in this paper should be interpreted as points of reflection, in the hope of contributing to and stimulating the policy debate. In no case should they be interpreted as final recommendations or absolute solutions.

I would like to thank Professor Gilles Grenier, Professor David Gray and my colleagues for their comments.

Abstract

There is a common belief that inequality can be good for economic growth when viewed as a reward for personal achievements due to merit, stimulating the drive for people to improve their skills for greater personal and monetary gain². On the other hand, inequality of opportunity is viewed as a negative consequence of unequal social and economic status within society, which is dampening economic growth³. Assuming this trade-off to be valid, it becomes important for policy decisions to qualify and quantify inequality sources accordingly. This allows us to identify the relative influence of inequality sources and to look at how structural and labour composition changes influenced the inequality variation by sources between points in time.

The data used are the variables of Survey of Consumer Finances 1987 and 1997, individual files. The focus is placed on education, gender and presence of children – upon which social policies can be and are based. The wage inequality decomposition using a step-wise procedure applied on a typical earning equation shows that education and gender are two significant sources of inequality, while the presence of children has a lesser impact. The structural changes for education increases inequality, while the reduction in the wage gender gap partially offset the increase. Overall, from 1987 to 1997, the wage inequality of individuals remained stable with a small decrease. The main policy recommendations are to focus on the equality of opportunity, especially in the education sector; meanwhile, current gender policies seem to be achieving their goals of reducing the gender gap. The paper shows the usefulness of an earning equation in inequality decomposition, but it also highlights the need for more research, both on the decomposition methodology and the overall policy development in its broader context.

² Bell and Freeman 2000.

³ Altman 2001.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

"Any city however small, is in fact divided into two, one the city of the poor, the other of the rich; these are at war with one another."

– *Plato, The Republic, 370 BC*

Introduction

The debate concerning 'poor' people and 'rich' people seems to have always been present in civilised society. This is a complex issue where some think that the rich are rewarded because they deserve more, implicitly stating that the poor do not deserve as much. Others think that the rich do not deserve as much as they have, and that therefore greater equality should be a social and economic objective. In recent years, especially with some setbacks of communist societies, equality at all costs does not seem as appealing as before, but there is still a general consent that the income gap between the poor and the rich should not be ignored. Inequality concerns, therefore, remain of interest for social policy development.

Addressing inequality issue without knowing its actual sources and effects can be a waste of resources giving rise to perverse results – creating disincentives to work and social exclusion. Decomposing inequality allows us to identify where the government could concentrate its resources. Identifying the sources of inequality can also give us an idea of the magnitude and time frame of the resources needed to reduce it. For example, it is generally accepted that education yields beneficial results for the whole society, although most of the benefits will be reaped in future years.

The scope of this research is to examine wage inequality sources using a typical earning equation and to draw plausible policy implications. The paper contains four

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

sections. The first reviews and discusses the inequality question, measurement and decomposition issues. The second section presents the inequality decomposition methodology developed by this author. The methodology is an intuitive approach where impact analysis is used to identify the inequality sources. Section 3 applies the methodology using two points in time, 1987 and 1997, and the population aged 25 to 64; from which an earning equation based on the human capital theory is developed. The equation is used to decompose the wage level, using a sequential step-wise procedure in conjunction with the Gini coefficient measure in order to identify the inequality sources. The fourth section analyses the research findings and their policy implications.

1 Inequality

2372 years ago, Plato was already paying attention to the difference between the poor and the rich. There are continuous debates about the meaning and importance of both inequality and poverty measures. Some argue that only absolute poverty measures should be considered, while others think that what is important is the relative poverty measures, where inequality plays a major role.¹ Andrew Coyne, a columnist for the National Post, expressed his disagreement with the whole debate about poverty measurements and concluded his article by saying (about the absolute and relative measures of poverty respectively): "One is relatively absolute. The other is absolutely

¹ For examples and definition of absolute and relative poverty, see Ruggles 1990.

relative."² This quote highlights the fact that as the society evolves, the definition and conception of poverty evolves as well. The debate is therefore unlikely to disappear, and it brings the need of a better understanding of the driving force behind relative poverty and, implicitly, inequality.

1.1 Inequality Measurement Issue

Inequality looks at the difference between poor and rich, but instead of focussing on the lower end of the distribution, it considers the overall distribution. Appendix I presents a brief list of existing inequality measures. One of the problems with most inequality measures is that they either require the absence of zeros or of negative values³ for wealth or income. Moreover, most inequality decomposition techniques, by which inequality is decomposed by contributing sources, can only compare one group to another, and the importance of a particular sub-grouping attribute will vary depending on the measure of inequality used⁴.

² National Post, July 27, 2001.

³ For example, by looking in Appendix I, we can see that General Entropy measures do not constrain the values but, as the mean tends toward zero, the index tends toward infinity... The Hoover coefficient is another measure with no constraint on the value as long as the total wealth is not zero. Even if these measures have interesting proprieties, they are not frequently used in socio-economic because they do not convey an intuitive graphical interpretation such as the Gini coefficient does.

⁴ For example, referring to Appendix I, let us presume that we applied a decomposition technique that yields two series $\{E_i/E_{tot}-A_i/A_{tot}\}$, $\{-1 -1 0 1 1\}$ and $\{-2 0 0 0 2\}$ for factor A and B respectively. Then, using the Hoover coefficient: $Z_{Hoover} = \sum_{i=1..N} (abs(E_i/E_{tot}-A_i/A_{tot}))/2$, both factors A and B account for 2

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

The measure used for this research is the Gini coefficient, because it presents a simple and intuitive graphical concept bounded from 0 to 1⁵, and it has the advantage of permitting the use of zero values. The Gini coefficient is also frequently used in socio-economic research and, therefore, is the best-known inequality indicator. The exact nature of the decomposition used with the Gini is presented later in Section 2. The methodology has the advantage of not requiring sub-grouping but, unfortunately, it does not avoid the sensibility problem to the inequality measure used⁶.

1.2 Previous studies on inequality sources

The literature on income inequality is fairly exhaustive. But decomposition of inequality by sources is a relatively new concept. We had to wait until the late seventies and early eighties before seeing more papers on inequality decomposition⁷. In the Canadian literature, Wolfson (1998) was among the first to raise the question of

points each. Now, using the Coulter coefficient: $Z_{\text{Coulter}} = \text{sqrt}(\sum_{i=1..N} ((E_i/E_{\text{tot}} - A_i/A_{\text{tot}})^2)/2)$, factor A accounts for 1.41 points while factor B accounts for 2 points. This simple example illustrates that for a specific decomposition technique, using different inequality indicator can yield different results. See Litchfield 1999 or Fields 2001 for more details.

⁵ See Appendix I.

⁶ Inequality indicators have different properties and the different speed to which indicators react on a contraction toward the mean will influence the decomposition. The methodology proposed in Section 2 does not attempt reconcile this aspect for any indicator.

⁷ For example, Shorrocks 1982.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

inequality sources. Using sub-grouping inequality and polarisation⁸, his study shows that the inequality of earnings for men has increased between 1974 and 1995, while the inequality of earnings for women has decreased⁹ due to higher labour force participation.

More recently, a study done at the Department of Finance by Joanis and Rodriguez (2001) used a sub-group approach to analyse inequality trends. Using both the Theil index and the Gini coefficient, the authors found that family market income inequality increased despite a relatively stable trend in wage inequality. They explain this by the structural change in employment (shift toward more self-employment, full-time/part-time, etc.) and in the demographic (age, single parent, number of children, etc.). They rightfully noted that the Canadian Tax and Transfer System acts as an automatic stabiliser offsetting market income inequality fluctuation. It can also be noted that they did not find regional disparities to be an important factor.

Fields (2001) developed a new method to use the information contained in income-generating equations to decompose the shares attributed to each explanatory factor. In an application to rising earnings inequality in the United States, he found that schooling is the single most explanatory variable and all of schooling's effect was a coefficient effect (structural effect) because the rate of return on education has risen.

⁸ Polarisation being a special case of inequality where a reduction of the number of people in the middle class is observed.

⁹ Beach and Slotsve (1996) had similar findings looking at inequality between 1972 and 1992.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

In January 2001, the IRPP/CSLS¹⁰ organised a conference titled "Linkages between Economic Growth & Inequality". Some of the concluding remarks¹¹ by William Watson from McGill University were highlighting the need to study further the inequality questions. Some of his questions and comments were very legitimate:

- How to promote equality for growth - need to look at details such as which policies/ programs would be the most successful?
- Governments should rectify market failures that the market is not equipped to address, ex. Education.
- Worry about incentives.
- Need a human capital strategy.
- Do we need more research and development?

To answer the first two selected points, inequality decomposition can clearly play a major role. First, it can identify where market failures occur. Secondly, the nature of the inequality itself can prescribe how policies can restrain the total earning share of the wealthy or help the less fortunate. Implicitly, this requires concern about the incentives and the need to consider the broader human capital strategy (needs and goals). It is also clear that the inequality debate can only benefit from more research and new development analysis tools.

¹⁰ Institute for Research on Public Policy and Centre for the Study of Living Standards

¹¹ See CPAC (2001).

2 Methodology – Decomposition by factor and time

Decomposing inequality is problematic for many reasons¹². Most decomposition techniques, for example, will yield different results depending on the inequality measure used. Moreover, most widely used inequality measures contain the concept of variance relative to the mean and are, therefore, mean dependant. This makes the inequality measurements dependent on location of the distribution. In addition, not all inequality measure can handle non-strictly positive values.

The methods employed here are sequential step-wise decompositions. While they do not provide a unique decomposition, the methods give an intuitive approach and set a consistent framework to obtain meaningful impact analysis. This is especially important since the results become meaningful only when compared to each other, and cannot be interpreted as absolute measurements.

The first question is: Using a typical earning equation explaining wages of the form

$$\ln Y = \beta_0 + \sum_{k=1}^K \beta_k X_k + \varepsilon, \quad (1)$$

how can we identify the contribution of each factor k to the overall wage inequality?

To answer this question, we will use a method assessing the impact of removing the portion of income explained by a factor and then, redistributing equally the aggregate amount to each unit.

¹² For more details, see Litchfield 1999 or Kluge 2000.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

The second question is: Estimating equation (1) at two points of time,

$$\ln Y_{,1} = \beta_{0,1} + \sum_{k=1}^K \beta_{k,1} X_{k,1} + \varepsilon_{,1}, \quad (1')$$

and

$$\ln Y_{,2} = \beta_{0,2} + \sum_{k=1}^K \beta_{k,2} X_{k,2} + \varepsilon_{,2}, \quad (1'')$$

where subscript ",1" and ",2" indicate points in time,

how much of the difference in income inequality can we assign to the attribute effect (labour composition effect) and to structural change (coefficient effect)?

To answer this second question, we use a method similar to the Oaxaca (1973) decomposition, which is also used and demonstrated in Fields (2001).

2.1 Decomposition by explanatory factor

The idea behind the decomposition technique developed here is fairly simple. We estimate the income portion explained by a specific variable. We remove that amount from each unit receiving it, and redistribute the total amount evenly to all units. This answers a question like: "What would happen to inequality if, all other things being equal, men and women were receiving the same wage?"

Let define $I(Y)$ as a Gini Coefficient applied on the vector of incomes $Y = (Y_1, Y_2, \dots, Y_n)$, n being the sample size.

Now under:

$$\ln Y = \beta_0 + \sum_{k=1}^K \beta_k X_k + \varepsilon \quad (1)$$

we observe the impact of going from

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

$$I(Y_1, Y_2, \dots, Y_n) \quad (2)$$

to

$$I(Y_1 \times Y_{1k}^{-1} + \mu_k, Y_2 \times Y_{2k}^{-1} + \mu_k, \dots, Y_n \times Y_{nk}^{-1} + \mu_k) \quad (3)$$

for each k ,

where $\ln Y_{ik} = \beta_k X_{ik}$ and $\mu_k = (\sum_{i=1}^n Y_i - Y_i \times Y_{ik}^{-1})/n$.

Let us note that we remove the explained part by multiplication, since

$$Y_i \times Y_{ik}^{-1} \equiv \exp(\ln Y_i - \ln Y_{ik}),$$

but we redistribute evenly by addition with

$$\exp(\ln Y_i - \ln Y_{ik}) + (\sum_{i=1}^n Y_i - Y_i \times Y_{ik}^{-1})/n$$

instead of using a proportional redistribution of the form of

$$\exp[\ln Y_i - \ln Y_{ik} + (\sum_{i=1}^n \ln Y_{ik})/n].$$

If we compare this last expression with equation (3), leaving in the term $(\sum_{i=1}^n \ln Y_{ik})/n$, the exponential would result in a multiplication by a constant, i.e., a proportional redistribution. In this inequality decomposition methodology, we want to redistribute evenly and we therefore use the addition operator as specified in equation (3).

The marginal impact of going from equation (2) to (3) for each factor will not add up to the initial inequality coefficient, because each factor reduces the variance and therefore changes the inequality measurement. The Gini coefficient does not provide a solution to this problem. The resulting implications, however, do not change, but it does require more attention at the interpretation stage. To get an additive decomposition, we repeat the same process looking at the impact of going from equation (2) to (3') below,

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

for $Z = 1, 2, \dots, K$; adding one explanatory factor at a time:

$$I(Y_1 \times \prod_{j=1}^Z Y_{1j}^{-1} + \mu_Z, Y_2 \times \prod_{j=1}^Z Y_{2j}^{-1} + \mu_Z, \dots, Y_n \times \prod_{j=1}^Z Y_{nj}^{-1} + \mu_Z), \quad (3')$$

where $\mu_Z = (\sum_{i=1}^n Y_i - Y_i \times \prod_{j=1}^Z Y_{ij}^{-1})/n$.

The results of the decomposition depend on the order in which variables are added.

2.2 Decomposition by attribute and structural effect over time

To explain the change between points in time, we decompose the variation using the same principle as the Oaxaca (1973) decomposition¹³. The traditional example for Oaxaca's decomposition is the following: using equation (1') and (1''),

$$E[\ln Y_{i,2}] - E[\ln Y_{i,1}] = X_{i,2} \beta_{,2} - X_{i,1} \beta_{,1} \quad (4)$$

$$= X_{i,2} \beta_{,2} - X_{i,1} \beta_{,2} + X_{i,1} \beta_{,2} - X_{i,1} \beta_{,1}$$

$$E[\ln Y_{i,2}] - E[\ln Y_{i,1}] = (X_{i,2} - X_{i,1})\beta_{,2} + X_{i,1} (\beta_{,2} - \beta_{,1}) \quad (4')$$

$$\text{Mean}[\ln Y_{i,2}] - \text{Mean}[\ln Y_{i,1}] = (\bar{X}_{,2} - \bar{X}_{,1})\beta_{,2} + \bar{X}_{,1}(\beta_{,2} - \beta_{,1}) \quad (4'')$$

Where the first term of the left-hand side of equation (4') is the attribute effect and the second term of the left-hand side is the structural difference. Oaxaca suggested that this decomposition be computed at the means (equation (4'')).

The differences between the Oaxaca decomposition and the technique used in this research is that instead of using the expectation operator to identify the effect at the mean,

¹³ See Greene 2000, pp. 251-252 for details.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

we use the Gini and the inequality decomposition presented in the previous section. This also requires using the complete distribution instead of looking the result at the mean. Fields (2001) used the same technique, stating that with the auxiliary equation

$$\ln Y_{,aux} = X_{,1}\beta_{,2}, \quad (5)$$

then

$$I(Y_{,2}) - I(Y_{,1}) = (I(Y_{,2}) - I(Y_{,aux})) + (I(Y_{,aux}) - I(Y_{,1})), \quad (6)$$

where $I(Y_{,2}) - I(Y_{,aux})$ is the attribute or labour composition effect and $I(Y_{,aux}) - I(Y_{,1})$ is the structural or coefficient effect. Similarly, the variation in the decomposition of the Gini coefficient (DG)¹⁴ for the kth factor is then:

$$DG_k[I(Y_{,2})] - DG_k[I(Y_{,1})] = \{DG_k[I(Y_{,2})] - DG_k[I(Y_{,aux})]\} + \{DG_k[I(Y_{,aux})] - DG_k[I(Y_{,1})]\} \quad (7)$$

where $\{DG_k[I(Y_{,2})] - DG_k[I(Y_{,aux})]\}$ is the attribute or labour composition effect explained by the factor k and $\{DG_k[I(Y_{,aux})] - DG_k[I(Y_{,1})]\}$ is the structural or coefficient effect of the kth factor. If the decomposition DG is the additive decomposition, then summing equation (7) for k=1 to K, gives back equation (6). As stated earlier, let us keep in mind that if the decomposition is the marginal effect, the marginal effect of each variable cannot be summed to the initial inequality coefficient.

¹⁴ This is done using the inequality decomposition techniques presented in section 2.1. Field (2001) uses his own decomposition technique but applies the same approach as equations (6) and (7).

3 Decomposition Application – Data, Earning Equation and Results

The methodology is applied to the Canadian situation using the Survey of Consumer Finances (SCF), which represents a rich source of Canadian microdata. The SCF is a cross-sectional survey that contains income as well as personal and labour-related characteristics of individuals aged 15 years and over. It allows therefore the development an earning equation and the application of the decomposition methodology.

3.1 The data

The files used are the SCF 1987 and 1997 Income¹⁵, Individual Public-Use Microdata File – 1997 being the last year for which SCF data exist. Note that the income variables are for income received respectively in 1987 and 1997 calendar year, but the

¹⁵ We know that the SCF 1997 file has a problem with its wage distribution when compared with Canada Customs and Revenue Agency's Personal Income Tax Statistical Master file. The SCF 1997 overestimate the total wage and salaries by 7% with some disparities by income range. We know the problem to be related to the weighting scheme; the correlation between the variables within each record remains therefore consistent. This is why no mandatory adjustment is required to the wage data, as we do not expect the weight bias to affect the results.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

other characteristics, such as demographic and labour-related variable, are as of the third week of April 1988 and 1998 respectively – when the surveys were conducted¹⁶.

There is no explicit action taken to control for possible business cycle effects aside from the cycle position of the 1987 and 1997 years, which can be argued to be relatively similar. The selected population is individuals aged 25 to 64. Individuals with self-employment income different than zero were excluded during the earning equation development¹⁷. No other restriction was applied to the sample. In particular, the individuals with no earnings were retained because they largely influence the degree of the overall inequality. Many social objectives surrounding the inequality questions aim at assisting those without sufficient income, including those without earnings.

Table 3.1.1 Descriptive Statistics SCF 1997

1997*	Number of Records	Weighted Population 25-64	Y (97\$) (Wages)	Average**		
				E (Education) (years)	XP (Experience) (years)	Hrs (Hours Worked)
Female	21,498	7,497,006	\$17,573	13.0	23.2	1,069
Male	19,351	7,057,059	\$33,094	13.2	22.8	1,544
With kids	18,154	6,082,982	\$27,282	13.5	18.9	1,386
No kids	22,695	8,471,083	\$23,532	12.8	26.0	1,237
Not Mar.	10,353	4,051,480	\$21,897	13.2	20.5	1,215
Married	30,496	10,502,585	\$26,334	13.1	24.0	1,332

* 1997 Income, April 1998 demographic.

** See Appendix II for a detailed description of the variables.

¹⁶ For more details, contact the Income and Housing Survey Section, Household Survey Division of Statistics Canada or consult the SCF microdata file documentation.

¹⁷ This decision is later explained in Section 3.2.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Table 3.1.2 Descriptive Statistics SCF 1987

1987*	Number of Records	Weighted Population 25-64	Y (97\$) (Wages)	Average**		
				E (Education) (years)	XP (Experience) (years)	Hrs (Hours Worked)
Female	27,156	6,484,508	\$15,059	11.6	23.7	973
Male	23,391	5,946,525	\$34,888	11.9	23.0	1,663
With kids	25,993	5,670,788	\$25,901	12.0	19.3	1,339
No kids	24,554	6,760,245	\$23,407	11.4	26.8	1,274
Not Mar.	10,937	3,066,009	\$21,804	12.0	20.7	1,286
Married	39,610	9,365,024	\$25,442	11.6	24.3	1,309

* 1987 Income, April 1988 demographic.

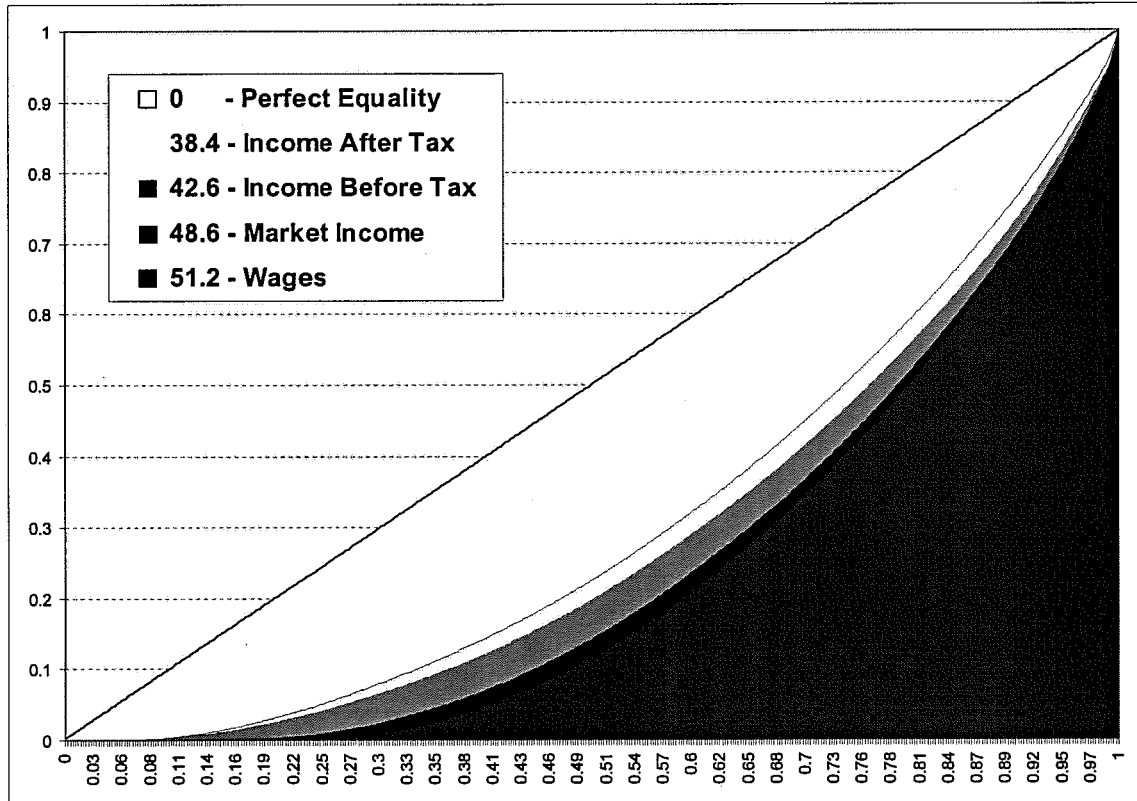
** See Appendix II for a detailed description of the variables.

Tables 3.1.1 and 3.1.2 present some descriptive statistics of the selected samples. One noticeable change is that males worked fewer hours in 1997 than in 1987, while the opposite trend applies to females. Any inequality study not controlling for the working hours could therefore be misleading¹⁸. The hours worked are certainly an explanatory factor of the fact that the average wages increased in constant dollar terms for women but not for men¹⁹. We can also observe that people became generally more educated between 1987 and 1997, and that there were fewer married individuals and fewer people living with children in 1997 than in 1987. These elements are to be kept in mind for Section 4 where the findings and the policy implications are discussed.

¹⁸ See Bell and Freeman 2000.

¹⁹ This could be a result related to the issue of the shift toward non-standard work. See Betcherman and Lowe 1997.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION



From the Survey of Consumer Finances 1997 income - Individual file
 Working age (25-64) population with self-employment equal to zero.
 The numbers in the legend are the Gini coefficients for each income source.
 Negative incomes are set to zero for the Gini calculation.

Figure 3.1 Income Inequality of the selected population 1997

To provide further background, Figure 3.1 presents Lorenz curves of the selected population in 1997 for different types of income. The X-axis is the cumulative population proportion and the Y-axis is the cumulative share of the total income. Drawing the Lorenz Curve for each income type, we can see that wages are the most unequally distributed with a Gini coefficient of 51.2%, followed by the market income²⁰ with a Gini

²⁰ Market income is the total income before tax without the government transfers. Consult the SCF microdata file documentation for the exact definition of total income before tax and government transfers.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

coefficient of 48.6%. Government transfers reduce the income before tax Gini coefficient to 42.6% and the progressiveness of the tax system reduces it further to 38.4%²¹.

3.2 The regression model for the earning equation

The model is based on the general human capital theory²², and the variables were chosen accordingly. Appendix II presents the definition of the variables and four different models tried through the development process of this study (see Table A2).

Models 1 and 2 include a self-employment income flag in an attempt to control for the disturbance introduced by the self-employed. Obviously, people with self-employment income are more likely to earn income by non-traditional means. The self-employed flag was significant, but the model improved with the removal of records with self-employment income different than zero. The flag was not sufficient to offset the unexplained variance introduced by those records. It is generally recognized in the policy field that the needs of non-traditional workers (mainly self-employed) are better analysed separately. Using the self-employed would add variance and complexity to the analysis without bringing additional information to the policy debate for traditional workers. To keep this study simpler and focussed, the records with self-employment different than zero were therefore removed from the sample in Models 3 and 4.

²¹ The same findings were highlighted in Joanis and Rodriguez 2001. This is a well-known fact.

²² For example, see Becker 1993.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Model 1 uses the wage, without the log, as the dependent variable. The use of the natural logarithm of the wage, from Model 2 and on, provided a better fit in terms of R-square, but it creates a problem if we want to keep the records reporting no earnings from wages. Removing the record without wage does not improve the fit. Moreover, it would create a major distortion in the inequality measure. To avoid this problem, the whole wage distribution was moved to the right by adding one dollar. This has no significant impact on the wage inequality²³ since adding a dollar to each record adds a total amount that is relatively small compared to the total wage earnings.

Similarly, using the log of hours worked improves the fit of the model, providing a linear relation in logs between wage and hours worked. From the model fit point of view, Model 4 improved over Model 3 to an R-squared of 65%, which is very high for these types of model. Model 4, reproduced in Table 3.2.1, has the best rationale²⁴ and the best fit of the models tried and is, therefore, the earning equation selected to apply the decomposition technique presented in Section 2.

²³ It does introduce a bias, however, in the regression coefficient because the usual assumptions on error terms do not hold. Other methods, such as Tobit or Heckit estimators (see Greene 2000, pp. 905-933 for details), could be used to estimate a consistent model. For simplicity reasons and to remain focused on the inequality decomposition aspect, these methods are not further investigated in this study.

²⁴ Linear relation between the log of wage and the log of hours worked and exclusion of the self-employed.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Table 3.2.1 Selected earning equation

Model 4 – 1997								
<i>Selection: Individual aged 25 to 64 with self-employment income = 0</i>								
$\ln Y =$	$\beta_0 +$	$\beta_1 E +$	$\beta_2 XP +$	$\beta_3 XP^2 +$	$\beta_4 \ln Hrs +$	$\beta_5 Mst +$	$\beta_6 nK +$	$\beta_7 M$
β_i values	2.1835	0.0345	0.0467	-0.0016	0.8778	0.2742	0.1790	0.9339
(t-stat)	(27.9)	(9.0)	(11.1)	(-19.7)	(226.8)	(8.7)	(5.9)	(36.4)
R^2	0.65						Number of Records: 40,849	
Model 4 - 1987								
β_i values	2.9294	0.0183	0.0257	-0.0013	0.8148	0.1604	0.1414	1.3332
(t-stat)	(40.3)	(5.3)	(6.5)	(-17.3)	(221.8)	(5.05)	(4.8)	(53.6)
R^2	0.61						Number of Records: 50,547	

See Appendix II for a detailed description of the variables.

As we can see in Table 3.2.1, Model 4 remained very good when used with the SCF 1987 instead of 1997 – keeping an R-square of 61%. All the coefficients of the models for both years are very significant with p-values lower than 0.1 %. The explanatory variables act as expected – education, experience (at a decreasing rate), hours worked, marital status and absence of children having positive effects on earnings.

3.3 Gini coefficient inequality decomposition by factor

Among the explanatory variables of the earning equation, the factors of interest as far as targeting possible interventions, from a social policy perspective, are education, gender and the presence of children. The other factors are less likely to be influenced by social policies. The number of hours worked depends mainly on labour market conditions and the policies surrounding this issue are unlikely to change. Labour market experience

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

is more closely related to the demographics and the marital behaviour in Canada. Again, these are unlikely to be affected by policies.²⁵

Table 3.3.1 Marginal effect on inequality of each factor, 1997

<i>SCF 1997</i>	<i>Gini Coefficients*</i>	<i>Point Variation</i>
<i>Wage</i>	51.17 %**	NA
<i>- Education</i>	31.79 %	19.38 %
<i>- Gender</i>	32.76 %	18.41 %
<i>- No kids</i>	47.53 %	3.64 %

See Appendix II for a detailed description of the variables.

* Gini coefficient after removing the variation explained by each and redistributing the effect evenly among all individuals, except for

** the wage where 51.17 % represent the initial inequality of wages in the sample without any modification.

Table 3.3.2 Marginal effect on inequality of each factor, 1987

<i>SCF 1987</i>	<i>Gini Coefficients*</i>	<i>Point Variation</i>
<i>Wage</i>	51.54 %**	NA
<i>- Education</i>	41.44 %	10.10 %
<i>- Gender</i>	29.37 %	22.17 %
<i>- No kids</i>	49.06 %	2.48 %

See Appendix II for a detailed description of the variables.

* Gini coefficient after removing the variation explained by each and redistributing the effect evenly among all individuals, except for

** the wage where 51.54 % represent the initial inequality of wages in the sample without any modification.

²⁵ As an example, Quebec decided to withdraw its Allowance for newborn essentially because the government realized that the policy did not had the desired impact on birth rates while there was better ways to help parents. Milligan (2001) indicates that the program had some success at increasing the fertility rate, but the demographic data clearly shows that it was not enough to offset the declining fertility trend. Without draconian measures, the social demographic behaviour will go its own way (for example, see Beaujot 2000), following cultural trends and not government policy objectives. We can note that policies related to the presence of children are not about inciting people to have more or less children. The policy objectives are to accommodate more or less the working parents by providing childcare programs or parental leaves. This is further discussed in Section 4.3.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Tables 3.3.1 and 3.3.2 present the marginal effect on the Gini coefficient of removing all the wages explained by a specific explanatory factor and redistributing the aggregate value evenly through the population as specified in equation (3). In 1997, we can see that removing the portion of wage explained by the education factor and redistributing the aggregate value to every individual, lowers the Gini coefficient by 19.38 points, from 51.17% to 31.79%. In 1987, education was not as influential on the inequality. The gender factor was clearly the most influential in 1987. It lowers the Gini coefficient by 22.17 points, from 51.55% to 29.38%. We can see in 1997 that education has a greater influence than gender or the presence of children. As explained in Section 2.1, the marginal effects are not cumulative and the results cannot be interpreted based only on the nominal numbers. These numbers give the relative importance of each factor at the margin within a fixed context.

Table 3.3.3 Gini coefficient additive inequality decomposition, 1997

<i>SCF 1997</i>	<i>Gini Coefficients</i>	<i>Point Variation</i>	<i>Cumulative Point Variation</i>	<i>% of Explained Inequality</i>
<i>Wage</i>	<i>51.17%</i>	<i>NA</i>	<i>NA</i>	<i>NA</i>
<i>- Education</i>	<i>31.79%</i>	<i>19.38%</i>	<i>19.38%</i>	<i>37.89%</i>
<i>- Education - Gender (male)</i>	<i>20.00%</i>	<i>11.80%</i>	<i>31.18%</i>	<i>60.92%</i>
<i>- Education - Gender (male) - No Children</i>	<i>18.34%</i>	<i>1.65%</i>	<i>32.83%</i>	<i>64.17%</i>
<i>- Education - Gender (male) - No Children - XP - Hrs - Mst</i>	<i>0.25%</i>	<i>18.10%</i>	<i>50.93%</i>	<i>99.52%</i>
<i>- Education - Gender (male) - No Children - XP - Hrs - Mst - Residuals</i>	<i>0 %</i>	<i>0.25%</i>	<i>51.17%</i>	<i>100%</i>

See Appendix II for a detailed description of the variables.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Table 3.3.4 Gini coefficient additive inequality decomposition, 1987

<i>SCF 1987</i>	<i>Gini Coefficients</i>	<i>Point Variation</i>	<i>Cumulative Point Variation</i>	<i>% of Explained Inequality</i>
<i>Wage</i>	51.54%	NA	NA	NA
<i>- Education</i>	41.44%	10.10%	10.10%	19.60%
<i>- Education - Gender (male)</i>	23.15%	18.29%	28.39%	55.08%
<i>- Education - Gender (male) - No Children</i>	21.63%	1.52%	29.91%	58.04%
<i>- Education - Gender (male) - No Children - XP - Hrs - Mst</i>	0.19%	21.45%	51.36%	99.64%
<i>- Education - Gender (male) - No Children - XP - Hrs - Mst - Residuals</i>	0 %	0.19%	51.54%	100%

See Appendix II for a detailed description of the variables.

Tables 3.3.3 and 3.3.4 present the additive inequality decompositions using the Gini coefficient as defined in equation (3') of Section 2.1. In this decomposition, the order in which each variable is taken matters and must be fixed for comparison between two points in time. The education line is equivalent to the marginal effect of the previous table because it is in the first position. The following line is the effect of removing the portion of wage explained by both the education and the gender factors, and redistributing the total aggregate value to every individual. The point variation is the effect on inequality of adding a subsequent factor or group of factors – for example, of going from looking at the education factor to considering both education and gender. This is repeated until the last line reduces the Gini coefficient to zero; removing all the wages, including the residual portion, and giving the mean to every individual, brings the Gini to zero by definition.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

The percentage of explained inequality is the cumulative point variation divided by the initial inequality (51.17 in Table 3.3.3). For example, using the additive decomposition with the earning equation in 1997 (Table 3.3.3) explained 99.52% of the total inequality or 50.93 points out of a total 51.17 points – the residuals accounting for the remaining 0.25. The additive decomposition is less interesting when comparing the factors to each other (because it is sensitive to the order of the factors in the decomposition), but it gives a better global view when comparing two points in time – as it is done in the next section.

3.4 Gini coefficient inequality decomposition by structure and attribute over time

To explain the change between two points in time, we decompose the variation with the method explained in Section 2.2. Using the SCF 87 and 97 to define two sets of wages and explanatory variables $\{Y_{,87}, X_{,87}\}$ and $\{Y_{,97}, X_{,97}\}$ of size n_{87} and n_{97} with the underlying Model 4 as defined in Section 3.2 and in Appendix II:

$$\ln Y_{i,87} = X_{i,87} \beta_{,87} + \varepsilon_{i,87}, \quad i = 1, \dots, n_{87} \quad (8)$$

$$\ln Y_{i,aux} = X_{i,87} \beta_{,97}, \quad i = 1, \dots, n_{87} \quad (9)$$

$$\ln Y_{i,97} = X_{i,97} \beta_{,97} + \varepsilon_{i,97}, \quad i = 1, \dots, n_{97} \quad (10)$$

the variation in the decomposition of Gini coefficient (DG) for the k^{th} factor is then:

$$\begin{aligned} DG_k[I(Y_{,97})] - DG_k[I(Y_{,87})] = \\ \{DG_k[I(Y_{,97})] - DG_k[I(Y_{,aux})]\} + \{DG_k[I(Y_{,aux})] - DG_k[I(Y_{,87})]\} \end{aligned} \quad (11)$$

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Reusing the results found in Tables 3.3.3 and 3.3.4, for 1997 and 1987, with the result of Table 3.4.1 for the set $\{Y_{,aux}\}$, the variation between the results of 1987 and of 1997 can be explained by taking the point variation difference between Tables 3.4.1 (result from the auxiliary model) and 3.3.4 (1987) to get the structure effect, and by taking the point variation difference between Tables 3.3.3 (1997) and 3.4.1 (result from the auxiliary model) to get the attribute effect. The total structural effect will always be zero by construction since we are looking at the underlying additive decomposition for the same year. The sum of the two effects gives the total variation.

Table 3.4.1 Gini coefficient additive inequality decomposition, auxiliary model equation (9)

SCF 1987 with β_{97}	Gini Coefficients	Point Variation	Cumulative Point Variation	% of Explained Inequality
Wage	51.54%	NA	NA	NA
- Education	34.36%	17.18%	17.18%	33.33%
- Education - Gender (male)	20.54%	13.82%	31.00%	60.15%
- Education - Gender (male) - No Children	18.92%	1.62%	32.62%	63.28%
- Education - Gender (male) - No Children - XP - Hrs - Mst	0.28%	18.64%	51.26%	99.45%
- Education - Gender (male) - No Children - XP - Hrs - Mst - Residuals	0 %	0.28%	51.54%	100%

See Appendix II for a detailed description of the variables.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Table 3.4.2 Additive decomposition by structure and attribute between 1987 and 1997

<i>SCF</i>	<i>Gini coefficient Point Variation</i>		
	<i>1987 with β_{87} to 1987 with β_{97} (Structure Effect)</i>	<i>1987 with β_{97} to 1997 with β_{97} (Attribute Effect)</i>	<i>1987 to 1997 (Total)</i>
Factors:			
<i>Education</i>	7.08%	2.20%	9.28%
<i>Gender_(male)</i>	-4.47%	-2.02%*	-6.49%
<i>No Children</i>	0.10%	0.03%	0.13%
<i>XP_(experience) + Hrs_(hours worked) + Mst_(married)</i>	-2.81%	-0.54%	-3.35%
<i>Residuals</i>	0.09%	-0.03%	0.06%
Total	0%	-0.37%	-0.37%

See Appendix II for a detailed description of the variables.

* This change is relatively big. It can probably be explained by the exclusion of the self-employed and be a result related to the shift toward non-standard work.

Table 3.4.2 presents the results where a positive number means that the effect in question increased inequality. For example, the biggest factor of increase is the structural change in education related to increased returns to education. Similarly, the biggest offsetting factor is the structural change in the gender effect. These results are further discussed in Section 4 below.

Note that in Table 3.4.2, the order of the factors in the additive decomposition affects the total Gini coefficient point variation between 1987 and 1997. This is a caveat of the additive decomposition presented in Section 2.2. On the other hand, the relative importance of the structure and attribute effect is independent of the order. The break down of the effects depends on the coefficients (the β s of the models) and the attributes themselves, which in both cases are obviously independent of the order. For example,

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

even if the education factor was not placed in the first position, the structure effect would remain the dominant factor over the attribute effect.

3.5 Results sensitivity

There are a few questions related to the sensitivity issue that are worth mentioning. First, the additive inequality decomposition is sensitive to the order of the variable. This is an important caveat and it requires an extra degree of caution when interpreting the results. The results make sense only when they are compared to each other or through time, and as long as the methodology remains the same. Under these conditions, the results are relatively robust. The model and its coefficients remain very significant independently of the year used. The variation follows a clear trend that is consistent with what we expect – closing gender gap reducing inequality, higher return to education increasing inequality, etc. (These results are further discussed in Section 4 below.) The different education codes, and recoding, in the two databases may introduce a small bias, but it is unfortunately difficult to estimate. The coefficients for education were nevertheless very significant for both years and it is not expected that the recoding bias the results.

4 Findings and Policy Implications

Decomposing inequality allows us to identify where the government could concentrate its resources. The results of the inequality decomposition can therefore be

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

used to feed in the policy debate. The final objective is to yield beneficial results for the whole society in an efficient way. This section reviews the findings and highlights the policy implications within the current Canadian context.

As explained in Section 3.3, the factors of interest from a social policy perspective are education, gender and the presence of children. The other factors (experience, hours worked and marital status) are, therefore, left outside of the focus of this study. Note that it was nevertheless important to include these factors in the earning equation model in order to improve the reliability of the analysis for the other factors of interest.

4.1 Education

Findings

The Canadian population was more educated in 1997 than 1987. As we can see by comparing the descriptive statistics in Tables 3.1.1 and 3.1.2 for 1997 and 1987 respectively, the education level increased on average by more than a year in each category. Due to the difference in education coding between the two samples, the increase may not be exact, but the trend is nevertheless significant. At the same time, the return on education was also higher in 1997 than in 1987 (see Table 3.2.1). Again, the exact level may be subject to a bias, but not sufficiently big to offset the trend.

When looking at Table 3.4.2, we can see that the changes in the education level (attribute effect) and in the return to education (structure effect) both contributed to an

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

increase in inequality²⁶. The structural effect contributed significantly more to the inequality increase than the attribute effect (7.1% Gini coefficient points for the structure effect compare to 2.2% for the attribute effect). The significant contribution to inequality increase of the structure effect was expected. Any increase in the education returns, with a log wage model, results in exponential wage increases and, therefore, in significant wage inequality increases.

The education attribute effect contributed in a much smaller way to the inequality increase. Note that even if the education attributes contributed to inequality, it does not mean that education inequality increased itself. If everybody increased their education level by 10%, the inequality will still go up because of the exponential return to education in the model. In fact, a quick calculation²⁷ of the education inequality indicates a sharp drop of the education Gini coefficient from 21.12% in 1987 to 16.59% in 1997 for the population aged 25 to 64 with self-employment income equal to zero. The final attribute effect remains, nevertheless, a positive factor in the inequality increase between 1987 and 1997 – simply because education pays just so much more!

Policy Implications

The inequality produced by the education factor is considered as inequality due to merit, which may stimulate economic growth. This type of inequality may stimulate the drive for people to improve their skills for greater personal and monetary gain. Clearly,

²⁶ This confirms the findings of Aghion and Howitt 2001.

²⁷ Calculated on our samples for 1987 and 1997 using the recoded education variables as defined in Appendix II.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

the governments in Canada do not want to limit the education of its citizens. Governments therefore have to concentrate their efforts at including education in the broader context of social inclusion. The key ideas are equality of opportunity. Governments can have a role in promoting education and at making it accessible to everyone. The federal government is already assuming a leadership in the education sector at the early childhood and at the post-secondary levels, and could even do more by increasing its investments in Early Childhood Development, transfers to provinces, Student Loan programs, Registered Education Saving Plan, Canada Education Saving Grants, and other policies.

4.2 Gender

Findings

There is a gender gap, but it is not the main scope of this research to identify the nature of the gap. The model controls for hours worked, experience, education, marital status and the presence of children. The model does not control for occupation or the industrial sector. So the nature of the gap could be due to personal choices, sectoral entrance barriers or pure discrimination. The good news is that this gap is closing and it contributed to a reduction in inequality.

Policy Implications

The gender gap can be viewed as a market failure that governments can address by regulations, awareness campaigns, education, etc. In recent years, the situation has improved and income inequality has been reduced as a result. The question "Can

governments do more and is it desirable?" is a highly politicised and normative question with no simple answer. Current policies and laws²⁸ seem to be working and should be maintained. More research would be required to efficiently answer the question as to whether more should be done. The nature of the gap has to be more explicitly investigated. Each underlying trend has to be monitored to detect if more policy interventions are required. At the federal level, this falls under the mandate, in collaboration with other government department, of the Status of Women Canada department.

4.3 Presence of children

Findings

By comparing Tables 3.1.1 and 3.1.2 "Descriptive Statistics" for 1997 and 1987, we can see that there were fewer individuals living with children in 1997 than 1987. The presence of children is a problem for working individuals – slightly more so in 1997 than in 1987 as we can see in Table 3.2.1. The presence of children was not a dominant factor in the inequality decomposition. In 1997, it had a first position effect of 3.64 Gini points compared to 18.41 for gender and 19.38 for education (Table 3.3.1). It still contributed to a small increase in inequality.

²⁸ Policies such as programs falling under the Gender Equity Agenda and laws such as the *Canadian Charter of Rights*, the *Federal Labours Code*, the *Employment Equity Act* and others.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

We intuitively know that the absence of children leaves more time and energy for work and, therefore, tends to result in higher wage. The positive correlation between the absence of children and the wage in our selected model confirms this. On the other hand, we know from the demographics that younger and older individuals have fewer children (young individual presumably because they did not have children yet and older individuals presumably because the children left home). At the same time, younger individuals have lower wages and older individuals have a declining return on experience as confirmed by the negative correlation in the selected model between the wage and experience squared. Although this would require further analysis, it is likely that these offsetting factors are also present in the inequality decomposition and contribute at explaining why the presence on children is not a dominant factor in the inequality decomposition.

Policy Implications

Policies to help workers with children, such as Daycare programs, Childcare Expense Deduction, Employment Insurance Maternity/Parental leave, etc., are rightly based on social values surrounding families with children. The point is not to incite a mother to go to work instead of staying at home with the children in an attempt to reduce inequality or poverty. The point is to give choices and opportunities to families. The effect of the presence or absence of children on inequality should, therefore, remain small and be based on the choices of individuals, and not on career barriers.

Conclusion

When used in a proper way, inequality analysis can be helpful in the policy formulation process. The addition of an earning equation in the inequality analysis can give a relative measure of source of inequality. It can also identify inequality trends resulting from the various explanatory factors. Further decomposing the trends in a structural and attribute effect provides even more useful input to policy formulation. While inequality due to merit is legitimate, the analysis allows identifying potential sectors where the government can intervene to provide equality of opportunity and to correct inequalities due to labour market failures. Good policies have to be consistent with the labour market demand in relation to a broader human capital strategy in order to maximise Canadian's position in the global economy.

Possible Future Steps

The debate surrounding inequality will likely continue to influence future policy development. Refining the earning equation and including other types of income could further improve the inequality analysis presented in this paper. This is likely to require different models, each adapted to the specific nature of the income source studied. The inequality decomposition methods should be further studied. The use of the Gini coefficient had some advantages, although it also brought some constraints. Other inequality measures could bring a different perspective – an avenue worth investigating. Finally, to maximise the usefulness of inequality analysis to policy development, the analysis should concentrate on specific policy questions in relation to the labour market and human capital strategy goals.

Bibliography

Aghion, P. and P. Howitt (2001), "Wage Inequality Within and Between Groups: A Schumpeterian Perspective", Brown University, Paper presented at the Centre for the Study of Living Standards conference on *Linkage Between Economic Growth and Inequality*, Ottawa, Canada, January.

Altman, M. (2001), "Economic Growth and Income Equality: Implications of a Behavioral Model of Economic Growth for Public Policy", Duke University, Paper presented at the Centre for the Study of Living Standards conference on *Linkage Between Economic Growth and Inequality*, Ottawa, Canada, January.

Battle, K. (2001), "Relentless Incrementalism: Deconstructing and Reconstructing Canadian Income Security Policy", in Banting, K., A. Sharpe and F. St-Hilaire (editors), *The Review of Economic Performance and Social Progress*, CSLS-IRPP.

Beach, C. and G. Slotsve (1996), *Are We Becoming two Societies?*, CD Howe Institute, Toronto.

Beach, C., R. Finnie and D. Gray (2001), "Earnings Instability of Women and Men in Canada 1982-1997", Queen's and Ottawa Universities, Paper presented at the Centre for the Study of Living Standards conference on *Linkage Between Economic Growth and Inequality*, Ottawa, Canada, January.

Beaujot, R. (2000), "La démographie et le Canada de demain : une réflexion sur les hypothèses de projection", University of Western Ontario, Paper presented at the Office of Superintendent of Financial Institutions' *Seminar on Demographic and Economic Perspectives of Canada Years 2000 to 2050*, Ottawa, Canada, March.

Becker, G.S. (1993), *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*, third edition, University of Chicago Press.

Bell, L. and R.B. Freeman (2000), "The Incentive for Working Hard, Explaining Hours Worked Differences in the U.S. and Germany", *NBER Working Paper*, No. 8051, National Bureau of Economic Research, Cambridge, MA.

Betcherman, G. and G. Lowe (1997), "The Future of Work in Canada", Canadian Policy Research Networks.

Bethlehem, J.G. and W.J. Keller (1987), "Linear Weighting of Sample Survey Data", *Journal of Official Statistics*, Vol. 3, No. 2, pp. 141-153, Statistics Sweden.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Brogan, D.J. (1998), "Pitfalls of Using Standard Statistical Software Packages for Sample Survey Data", in Armitage, P. and T. Collon (editors), *Encyclopedia of Biostatistics*, in section titled "Design of Experiments and Sample Surveys", John Wiley Publisher.

Canada's Political Channel (CPAC) (2001), *Economic Growth and Income Inequality*, Taped recording of the Centre for the Study of Living Standards conference on *Linkage Between Economic Growth and Inequality*, Ottawa, Canada, January 2001, last broadcast (E.S.T.): 2/22/2001 4:00:00 AM.

Card, D. and R.B. Freeman (1993), *Small Differences that Matter: Labor Market and Income Maintenance in Canada and the United States*, University of Chicago Press.

Coyne, A. (2001), "A tisket, a tasket, a basic needs basket", *National Post*, Canada, July 27, 2001.

D'Ambrosio, C. and C. Gradin, (2000), "Are Children in Growing Danger of Social Exclusion? Evidence from Italy and Spain", *Luxembourg Income Study Working Paper*, No. 262.

Fields, G.S. (2001), "Accounting for Income Inequality and its Change: A New Method, with Application to the Distribution of Earnings in the United States", Cornell University, School of Industrial Relation, Ithaca, NY 14853-3901, USA, July.

Greene, W.H. (2000), *Econometric Analysis*, fourth edition, Upper Saddle River, N.J., Prentice Hall.

Jackson, A. (2000), "Why We Don't Have to Choose between Social Justice and Economic Growth: The myth of the equity/efficiency trade-off", Canadian Council on Social Development.

Joanis, M. and E. Rodriguez (2001), "The Contribution of Labour Markets, Demographics and Public Redistribution to Inequality: A Decomposition Analysis for Canada, 1986-1996", *Department of Finance Working Paper*, 2001-18.

Kluge, G. (2000), "Entropy and Inequality measures", <http://poorcity.richcity.org> .

Kennickell, A.B., M. Starr-McCluer and B.J. Surette (2000) "Recent Changes in U.S. Family Finances: Results from the 1998 Survey of Consumer Finances", *Federal Reserve Bulletin*, January.

Litchfield, J.A. (1999), "Inequality: Methods and Tools", Text for the World Bank's web site on Inequality, Poverty, and Socio-economic Performance, <http://www.worldbank.org/poverty/inequal/index.htm> .

Milligan, K. (2001), "Subsidizing the Stork: New Evidence on Tax Incentive and Fertility", University of Toronto, Research paper for Social Science and Humanities Research Council and the Canadian International Labour Network.

DECOMPOSITION OF INCOME INEQUALITY USING AN EARNING EQUATION

Oaxaca, R.L. (1973) "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

Osberg, L. (2001), "Labour Supply and Inequality Trends in the United States and Elsewhere", Dalhousie University, Paper presented at the Centre for the Study of Living Standards conference on *Linkage Between Economic Growth and Inequality*, Ottawa, Canada, January, (check <http://is.dal.ca/~osberg/home.html> for most recent version).

Osberg, L. and A. Sharpe (2000), "International Comparisons of Trends in Economic Well-Being", *Luxembourg Income Study Working Paper*, No. 242.

Picot, G. (1998), "What Is Happening to Earnings Inequality and Youth Wages in the 1990s", Statistics Canada.

Ruggles, P. (1990), *Drawing the Line: Alternative Poverty Measures and their Implication for Public Policy*, The Urban Institute Press, Washington, D.C.

Smith, T.M.F. (1999), "Defining parameters of interest in longitudinal studies and some implications for design", University of Southampton, Department of Mathematics, Southampton, SO17 1BJ, U.K., Paper presented at the *International Statistical Institute conference*, Helsinki, Finland.

Shorrocks, A.F. (1982), "Inequality Decomposition by Factor Components" *Econometrica*, No. 50, pp. 193-212.

Wolfson, M.C. (1994), "When Inequalities Diverge", *American Economic Review*, Vol. 84, No. 2, pp. 353-358.

Wolfson, M.C. and B. Murphy (1998), "New views of inequality trends in Canada and the United States", *Monthly Labor Review*, Vol. 121, No. 4, pp. 3-23.

Appendix I Review of Inequality Measures

Table A1 Examples of inequality measure

General Entropy
index

$$GE(\alpha) = \{[\sum_{i=1..N}(E_i / E_{avg})^\alpha] / N - 1\} / (\alpha^2 - \alpha); \alpha \geq 0$$

Atkinson = Demand
index coefficient

$$Z_{Demand} = 1 - \exp(\sum_{i=1..N}(E_i \times \ln(A_i/E_i))/E_{tot}) \times E_{tot}/A_{tot} = 1 - Z_{MacRae}$$

Theil
redundancy

$$R_{Theil} = -\ln(1 - Z_{Demand}) = -\ln(Z_{MacRae}) \\ = \ln(A_{tot}/E_{tot}) + \sum_{i=1..N}(E_i \times \ln(E_i/A_i))/E_{tot}$$

Reserve
coefficient

$$Z_{Reserve} = 1 - \exp(\sum_{i=1..N}(A_i \times \ln(E_i/A_i))/A_{tot}) \times A_{tot}/E_{tot}$$

D&R
coefficient

$$Z_{D\&R} = 1 - \exp(-R_{KL}) = 1 - \sqrt{(1 - Z_{Demand}) \times (1 - Z_{Reserve})} \\ = 1 - \exp(\sum_{i=1..N}(\ln(A_i/E_i) \times (E_i/E_{tot} - A_i/A_{tot}))/2)$$

Kullback-Liebler
redundancy

$$R_{KL} = R_{D\&R} = -\ln(1 - Z_{D\&R}) \\ = \sum_{i=1..N}(\ln(E_i/A_i) \times (E_i/E_{tot} - A_i/A_{tot}))/2$$

Hoover
coefficient

$$Z_{Hoover} = \sum_{i=1..N}(\text{abs}(E_i/E_{tot} - A_i/A_{tot}))/2$$

Coulter
coefficient

$$Z_{Coulter} = \sqrt{\sum_{i=1..N}((E_i/E_{tot} - A_i/A_{tot})^2)/2}$$

Gini
coefficient

sort data: $E_i/A_i > E_{i-1}/A_{i-1}$

$$Z_{Gini} = 1 - \sum_{i=1..N}((2 \times \sum_{k=1..i}(E_k) - E_i) \times A_i) / (E_{tot} \times A_{tot})$$

• Variables:

- A_i : "people" (amount of individuals in group_i of a society), $A_{tot} = \sum_{i=1..N}(A_i)$
- E_i : "wealth" (total wealth owned by that group_i of a society), $E_{tot} = \sum_{i=1..N}(E_i)$
- N: amount of groups (quantiles, percentiles) in the society
- Z...: inequality measure for society (unified group, all groups)
- R...: redundancy (maximum entropy of society less actual entropy of society)

The measure used for this research is the Gini coefficient. Gini coefficients are frequently used in socio-economic research because they present a simple and intuitive graphical concept bounded from 0 to 1 presented in Figure A1. The X-axis is the cumulative population proportion and the Y-axis is the cumulative share of the total income (in general terms – the cumulative density function). The resulting curve is called the Lorenz Curve and the Gini coefficient is the ratio of the area between the 45 degree line (perfect equality) and the Lorenz Curve, divided by the total area under the 45 degree line. The Gini coefficient of a perfect equality is equal to zero and, at the opposite, is one for perfect inequality (where everyone is at zero except for one person who has everything).

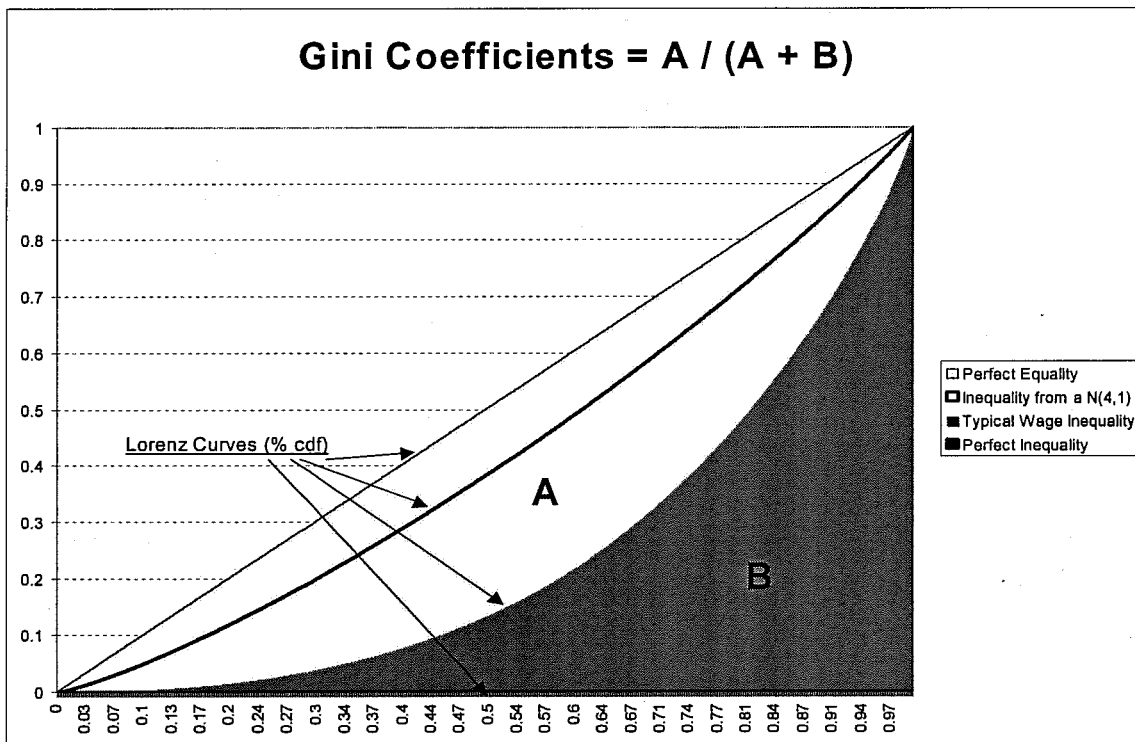


Figure A1 Gini coefficient calculation

Appendix II Earning Equation Model Definition

The microdata file used for constructing the model is the Survey of Consumer Finances (SCF) 1997 Income unadjusted, Individual Public-Use Microdata File, selected individuals aged 25 to 64. Note that in the models, the residuals are not explicitly stated but all the models have the form $Y = \text{Model} + \text{errors}$. The weights were not used¹. The dollars were all set in 1997 constant dollar using the 1992 based CPI from CANSIM series P200000.

The variables used are:

Dependant Variable:

Wages and Salaries	(Y)
--------------------	-----

Explanatory Variables:

Education (in years)	(E)
Experience (in years)	(XP)
Hours worked in the year	(Hrs)
Presence of Self-Employment Income	(SF)
Marital Status	(Mst)
Absence of Children	(nK)
Sex (male)	(M)

¹ The used of weights from survey design is risky. There is a huge amount of literature on the subject. As an example, see Smith 1999. For more details, see Brogan (1998).

Modification-Construction of variables:

Wages and Salaries

$$\ln Y = \ln (Y+1)$$

Education² 1997

Grade 8 or lower	E=4
Grade 9 - 10	E=9
Grade 11 - 13, did not graduate from high school	E=11
Grade 11 - 13, did graduate from high school	E=12
Some post-secondary education, no degree, certificate or diploma	E=13
Trade certificate or diploma from a vocational school or apprenticeship training	E=14
Non-university certificate or diploma from a community college, CEGEP, School of Nursing, etc.	E=15
University certificate below bachelor's level	E=16
Bachelor's degree	E=18
University degree or certificate above bachelor's level	E=21

Education¹ 1987

No schooling or elementary	E=4
Grade 9 - 10	E=9
Grade 11	E=10.5
Grade 12	E=11.5
Grade 13	E=12.5
Some post-secondary education	E=13
Post-secondary certificate or diploma	E=15
University degree	E=19

² The education codes are transformed in back into numbers of years using the author best judgment. As discussed in Section 3.5 this could influence the results but it is not expected to be significant enough to change the policy implication.

Experience (48 records with negative value were set to zero for 1997)

$$XP = \text{Max}(0, \text{Age} - E)$$

Hours worked in the year

$$\text{Hrs} = (\text{Total Hours Worked per Week}) \times (\text{Weeks worked in Reference Year})$$

$$\ln \text{Hrs} = \ln (\text{Hrs} + 1)$$

Presence of Self-Employment Income

$$SF = \text{Self-Employment Income} \neq 0$$

Marital Status

Married or living Common law

Mst=1

Other

Mst=0

Presence of Children under 18 years in the census family

Yes

nK=0

No

nK=1

Sex

Female

M=0

Male

M=1

Table A2 Regression models tried (1997) and selected model (4) for 1997 and 1987

	Model				
	β_i values				
	(t-stat)				
	1997				1987
	1	2	3	4	4
Dependant variable					
Y (Wage)	Yes	NA	NA	NA	NA
In Y (Log of wage)	NA	Yes	Yes	Yes	Yes
Independent variables					
β_0 (Intercept)	-19,115 (-35.7)	3.4397 (36.7)	3.1798 (36.6)	2.1835 (27.9)	2.9294 (40.3)
E (Education)	1,329 (51.2)	0.0744 (16.4)	0.0505 (11.8)	0.0345 (9.0)	0.0183 (5.3)
XP (Experience)	513.9 (17.8)	0.0434 (8.6)	0.0476 (10.2)	0.0467 (11.1)	0.0257 (6.5)
XP² (Experience squared)	-6.95 (-12.4)	-0.0018 (-19.0)	-0.0019 (-20.9)	-0.0016 (-19.7)	-0.0013 (-17.3)
Hrs (Hours worked)	11.10 (121.1)	0.0024 (148.1)	0.0028 (183.3)	NA	NA
In Hrs (Log of hours worked)	NA	NA	NA	0.8778 (226.8)	0.8148 (221.8)
SF (Self-employment flag)	-24,047 (-89.3)	-6.1453 (-130.3)	NA	NA	NA
Mst (Married or not)	1,578 (7.3)	0.3792 (10.0)	0.3397 (9.7)	0.2742 (8.7)	0.1604 (5.05)
nK (no Children)	-1,847 (-8.9)	0.1811 (5.0)	0.1947 (5.8)	0.1790 (5.9)	0.1414 (4.8)
M (Male)	9,046 (50.5)	0.4931 (15.7)	0.5378 (18.5)	0.9339 (36.4)	1.3332 (53.6)
Selection:					
Individuals aged 25 to 64	Yes	Yes	Yes	Yes	Yes
Self-employed	Yes	Yes	No	No	No
Number of records	46,565	46,565	40,849	40,849	50,547
R²	0.42	0.49	0.56	0.65	0.61