SOURCES OF INEQUALITY IN CANADA

By

Ian Rongve

B.Com. (Economics) University of Saskatchewan
M. A. (Economics) University of British Columbia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
Doctor of Philosophy

in
THE FACULTY OF GRADUATE STUDIES
ECONOMICS

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

1994
© Ian Rongve, 1994
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Economics

The University of British Columbia
Vancouver, Canada

Date Oct 2, 1994
Abstract

This thesis first presents a general procedure for decomposing income inequality measures by income source. The first method draws on the literature of ethical social index numbers to construct a decomposition based on a weighted sum of the inequality indices for the respective component distributions. The second method is based on the Shapley value of transferable utility cooperative games. The ethical and technical properties of the decompositions are examined, showing that the interactive technique has some previously known decompositions as special cases.

In the third chapter I examine the contribution of differences in educational attainment to earnings inequality using the interactive decomposition by factor sources, introduced in chapter two, of the Atkinson-Kolm-Sen inequality index. I first use an estimated sample-selection model to decompose predicted labour earnings of a random sample of Canadians into a base level and a part due to returns to education. I do this decomposition once ignoring the effect education has on the probability of being employed and once accounting for this fact. I then calculate the contribution of these two sources of earnings to inequality measured by a S-Gini index of relative inequality for the full sample as well as two separate age cohorts. The results indicate that approximately one half to two thirds of measured inequality can be directly attributed to returns to education while the interaction between the two sources post-secondary.

The fourth chapter uses the earnings model from the third chapter to conduct policy simulations for broadly based policies, low targeted policies, and high targeted policies. I demonstrate that the policies targeting low education individuals produce a larger increase in social welfare than do the other two types of policy.
# Table of Contents

**Abstract**

**List of Tables**

**Acknowledgement**

**1 Introduction**

**2 Decomposing Inequality Indices by Income Source**

2.1 Introduction. .............................................. 9

2.2 Ethical Inequality Indices. ................................. 10

2.3 Previous Decomposition Attempts. ......................... 13

2.4 Decomposing Inequality Indices .......................... 17

2.4.1 Interactive Decompositions .............................. 18

2.4.2 Shapley Decomposition ................................ 26

2.5 A Comparison of the Decompositions ..................... 35

2.5.1 Interactive verses Shapley Decompositions .......... 35

2.5.2 Old vs New Decomposition Methods .................. 39

2.6 Potential Applications. ................................... 44

2.7 Conclusion ................................................ 46

**3 Human Capital Models of Income Distribution**

3.1 How Education Affects Earnings .......................... 49

3.2 General Equilibrium Issues ............................... 52
List of Tables

2.1 Amount of Measured Inequality Attributable to Different Sources ..... 41
4.2 Estimates of the Earnings Equation. .......................................... 83
4.3 Decomposition of $I(y)$ for Earnings Decomposition 2 .................. 91
4.4 Amount of Measured Inequality Attributable to Different Sources ....... 93
4.5 Percent of Measured Inequality Attributable to Different Sources ....... 94
4.6 Marginal Decomposition for the OLS Model .................................. 96
4.7 Decomposition of $I(y)$ for Earnings Decomposition .................. 97
4.8 Amount of Measured Inequality Attributable to Different Sources ....... 98
4.9 Percent of Measured Inequality Attributable to Different Sources ....... 99
4.10 Marginal Decomposition for the Tobit Model ............................ 101
4.11 Decomposition of $I(y)$ for Earnings Decomposition .................. 102
4.12 Amount of Measured Inequality Attributable to Different Sources ....... 103
4.13 Percent of Measured Inequality Attributable to Different Sources ....... 104
4.14 Marginal Effects of Income Sources on Inequality .................... 105
4.15 Comparison of Decomposition Measures ................................. 107
4.16 Percent of Inequality by Shapley Decomposition ........................ 109
5.17 Government Expenditures on Education .................................... 116
5.18 Costs of Education (1985-1986) ........................................... 117
5.19 Cost of a Change in Education Level .................................... 117
5.20 Effect of Policies on Lifetime Inequality and Welfare .................. 120
5.21 Calculation of Cost of low Targeted Policy .............................. 121
5.22 Effect of Policies on Lifetime Inequality and Welfare, Discounted at 6% 122
5.23 Effect of Policies on Lifetime Inequality and Welfare, Discounted at 3% 124
5.24 Effect of Policies on Lifetime Inequality and Welfare, Discounted at 6% 125
5.25 Effect of Policies on Lifetime Inequality and Welfare 126
5.26 Effect of Policies on Lifetime Inequality and Welfare, Discounted at 6% 127

A.27 Data Collected 140
A.28 Cuts to the Sample 140
A.29 Sample Means of the Data 141

B.30 Estimates of the Taubman Equation 146
B.31 Estimates of the Layard and Zabalza Equation 146
B.32 Estimates of the Generalized LZ Equation 147
B.33 Estimates of the Earnings Equation 149
B.34 Estimates of the Restricted Earnings Equation 150
Acknowledgement

I would like to thank the various members of my thesis committee, David Donaldson, David Green, Bill Schworm, and Guofu Tan for their help and encouragement. I'd also like to thank Herb Emery, Ann Holmes, Cherie Metcalf, Henry Thille, Wayne Thomas, Shaun Vahey, Liz Wakerly, and the rest of the UBC grad students for discussions and general silliness that made grad school relatively enjoyable. Thanks also to UBC, SSHRCC and Mom and Dad for financial support.
Chapter 1

Introduction
It has been demonstrated time and again by empirical economists that there is a generally positive correlation between the level of education of an individual and his or her labour earnings. There are several competing theoretical explanations for this phenomenon including the human capital approach of Becker (1964) and the market signalling approach of Spence (1973). This means that differences in educational attainment translate directly into differences in earnings and thus suggests that one possible method of reducing earnings inequality is to reduce the differences in educational attainment. The degree to which this is a reasonable policy depend to great extent on the actual quantitative effect of education on earnings inequality. This thesis is concerned with determining the contribution of education to total earnings inequality in Canada. The primary assumption behind the empirical work is the human capital hypothesis. The general format that I follow is to first determine a reasonable method of allocating contributions to income inequality to different sources of income and then to use these results in an empirical examination of how inequality is affected by returns to education.

My interest in this topic derives from my overall interest in welfare economics. One of the driving forces in welfare economics is how to increase the welfare of a society. The usual method of doing this is to increase the aggregate income of society but there is another way in which an increase in welfare can be achieved and that is to ensure that the existing income is better allocated, in whatever way is considered better.

In this thesis I provide an answer to the question of how does differences in educational attainment amongst individuals affect the measured inequality. As a preliminary to that question two other questions must be asked. The first is how do we measure inequality and the second being how best to allocate the effects on overall inequality to individual sources of income.

There are typically two approaches to the measurement of inequality. The first method is the purely descriptive, or statistical approach to measuring inequality. In
this approach an attempt is made to describe the underlying distribution function that generates the observed income inequality, with no attempt to decide whether or not the observed income distribution is good or bad. The second method is the ethical approach, typified by the analysis in Blackorby and Donaldson (1978) and Chakravarty (1990). In the ethical approach to the measurement of inequality an attempt is made to quantify the social desirability of the existing distribution in relation to other possible distributions which have the same total income\textsuperscript{1}. This approach to inequality measurement demands more of the researcher as it requires a formalization of the societal preferences over income distributions and in fact there is a one-to-one correspondence between a given ethical inequality index and a given formalization of society’s preference relation over income distributions. In this thesis I take an ethical approach to the measurement of inequality throughout. This does not mean that the results themselves depend on this ethical interpretation, only the interpretation of these results depends on this type of inequality measurement.

Once the method of measuring inequality has been decided, the second question is how this measured inequality is to be allocated amongst different sources of income. Several methods of doing this exist in the literature, such as the Pseudo-Gini decomposition of Fei, Ranis and Kuo (1978), and the technique introduced in Shorrocks (1982), but these suffer to varying degrees from several problems such as discontinuity in specific areas of the income space, or intuitively unappealing results\textsuperscript{2}. Therefore the initial item on the agenda is a general discussion of the theoretical decomposition of ethical income inequality indices by income source. This analysis is presented in Chapter two of the thesis.

\textsuperscript{1}Sen (1992) takes this one step further by claiming that the ethical approach does not measure inequality at all, what it measures is the social \textit{badness} of a given income distribution.

\textsuperscript{2}Chapter two considers these problems in greater detail.
income inequality indices by income source and outlining what I consider to be the undesirable features of these decompositions. I then introduce three general procedures for decomposing ethical inequality indices. The first general method produces what is called a direct interactive decomposition because the decomposition method produces a term which measures the extent to which inequality within the various sources of income is counteracted by inequality between sources of income. For example consider the case of two sources of income, where source one and source two are negatively correlated. In this case an individual with a high income from source one will have a low income from source two. In this case the interaction term in the interactive decomposition will be negative, indicating that inequality in one source of income counteracts inequality in the other source of income. This decomposition answers questions of the form ‘How much of observed inequality is a result of income source $j$?’

The second method of decomposing inequality indices produces a marginal interactive decomposition. In this decomposition I consider how measured inequality would change if I removed a given income source from aggregate income. The resulting value is the marginal contribution of an income source to overall inequality. This decomposition also has a term which may be interpreted as an interaction effect. In this case it provides a base level of inequality from which the marginal effects are measured. This decomposition is useful in answering questions like ‘How much would inequality change if we removed income source $j$?’

The last general procedure for decomposing inequality indices is derived from the Shapley value of transferable utility games. With this decomposition I first decompose the welfare measure using the Shapley value and then use the decomposed welfare to decompose the inequality index. Unlike the previous decompositions, the Shapley method allows an exact allocation of inequality to individual income sources, with no need for an interaction term. Thus this decomposition is able to make statements like, ‘The total
effect of income source $j$ on inequality is $z$.

After introducing the three methods of decomposing inequality indices I compare the new methods amongst themselves, then compare these methods to the previously known methods. I demonstrate, through the use of examples, the ethical and technical properties of the new decompositions. I then conclude the second chapter with a discussion of some of the possible applications of the decomposition techniques developed.

The third chapter is a theoretical discussion of the possibility of a human capital approach to earnings inequality. I first discuss the main competing explanation for the correlation between education and earnings, the signalling hypothesis, and state why my approach requires the human capital approach. I then construct a simple theoretical model which allows me to determine the effect of general equilibrium considerations on inequality. I maintain that even should the simple model be inadequate, and it likely is inadequate, the exercise in the rest of the thesis is still a valuable one to do.

The fourth chapter builds directly on the material presented in the second chapter. It is an application to a specific problem of the decompositions in the second chapter. The major question that I ask is ‘What is the contribution of returns to education to earnings inequality in Canada?’

To appropriately answer this question I first must decompose labour earnings into a part due to returns to education and a part due to other personal characteristics. Labour economics has identified two primary ways in which education can affect the returns to education; the first is the human capital approach, typified by Becker (1964), where education is seen as actually increasing the productivity of the worker, and the second is the market signalling approach of Spence (1973), where education is acquired to signal to potential employers a high productivity person. The technique that I use to decompose earnings themselves is based on the human capital model and is not appropriate for a market signalling model.
The first step in determining the impact of education on inequality is to estimate an empirical earnings equation. Therefore in chapter three I begin by drawing on the empirical labour economics literature to estimate an empirical earnings equation, using education as one of the explanatory variables, for a cross section of individuals in Canada for the year 1986. All subsequent analysis uses this sample.

With the estimated earnings equation I proceed to decompose earnings, into a part due to returns to education and a part due to other factors, by conducting a series of counterfactual experiments. I consider what would happen to the observed earnings distribution if, instead of their actual education level, the individuals had a counterfactual education level and earnings were generated by the same earnings process. The counterfactual level of education varies but is usually the lowest education level possible, less than nine years of schooling. I then consider the returns to education to be the difference between the actual earnings level and the constructed counterfactual earnings level. I then have two vectors, one a return to education and the other a return to all other personal factors. An appendix considers the possible impact of general equilibrium considerations on the inequality analysis.

With the decomposition of earnings just described I then use the direct interactive decomposition of the S-Gini inequality index to evaluate the effect of the various sources of income on overall measured inequality. I compare the results for several cohorts, those less than sixty-five years old, those between thirty and forty years old, and those between fifty and sixty years old. I also examine two ways of estimating the earnings equation, straightforward Ordinary Least Squares, and a tobit specification that takes into account the response of the probability of being employed to different education levels. The results indicate that approximately one third to one half of inequality is due to education. Finally for the most general earnings decomposition, I consider how the results change when the marginal interactive and the Shapley decomposition of the S-Gini index are
used instead of the direct interactive decomposition.

The fifth chapter presents some experiments to determine the effect of various education policies on the contribution of education to earnings inequality. I concentrate on three types of policies; these are broadly based policies, policies targeted at low education individuals, and policies targeted at high education individuals. In the fifth chapter I consider only the effect of the policies and not the specific form that they take.

The policies are examined in terms of their effect on the distribution of lifetime earnings and its effect on both inequality and social welfare. To do the lifetime analysis I use the earnings generating model developed in chapter four.

Two cases are considered for each policy; the case where the number of individuals affected by each policy remains constant and the cost of implementing the policy is allowed to vary, and the case where the cost of implementation is constant and the number of individuals affected varies. I show that, in both cases, the policy targeting low education individuals keeps earnings inequality approximately the same but does increase social welfare. The final chapter concludes.
Chapter 2

Decomposing Inequality Indices by Income Source
2.1 Introduction.

Most people obtain income from a variety of sources, such as returns to education, property income or government transfer payments. This fact is often not considered in empirical studies of income inequality that consider only the distribution of total income. At the same time these studies are frequently cited as justification for economic and social policies which attempt to decrease income inequality by targeting one particular source of income. In examining the effectiveness of such policies, a researcher needs a measure of income inequality that can be decomposed into the amount of inequality attributable to different sources of income.

Many papers have attempted to construct inequality indices that can be decomposed into a weighted sum of sub-indices, defined over the different factor components of income. Attempts of this sort include Fei, Ranis and Kuo (1978), Shorrocks (1982) and West and Theil (1991). In this chapter I first examine the previous theoretical attempts to perform decompositions and then introduce two different methods of performing the decomposition for arbitrary numbers of income sources. The new methods are both intuitively appealing and empirically tractable. Two of the methods that I introduce explicitly recognize that, for example, in an economy with two types of income, there are three potential sources of income inequality, income source one, income source two and the interaction between the income sources. The third method successfully allocates this interaction effect to the individual income sources.

I adopt an explicitly ethical approach to decomposing income inequality indices and begin by decomposing an equally distributed equivalent income measure for a given income distribution. Then the decomposed equally distributed equivalent income is used to construct a decomposition of the Atkinson-Kolm-Sen (AKS) family of inequality indices and the Kolm-Blackorby-Donaldson (KBD) family of inequality indices. I show how the
Chapter 2. Decomposing Inequality Indices by Income Source

ethical approach to decomposition allows for an easy and intuitive interpretation of the terms in the decomposition.

I compare the new decompositions presented here with some previously known decompositions and present in detail the differences between my decompositions and previously known decompositions. Then it is demonstrated that two of the previously known decompositions, that of the Theil inequality index and the standard decomposition of the variance of income, can be considered special cases of one of my decompositions for the AKS inequality index.

The rest of the chapter is as follows. In section two I introduce the basic notation and the ethical inequality indices that will be used. Section three discusses the previous literature on the decomposition of inequality indices by income source. The interactive and Shapley decompositions of the inequality indices are introduced in section four and in section five they are compared with each other and the previously known methods. Section six contains a discussion of some potential applications of the techniques and section seven concludes.

2.2 Ethical Inequality Indices.

An index number is said to be ethical if the value judgements inherent in the index are made explicit in the formulation of the index. In this section of the paper, I briefly explain the main results in the theory of ethical income inequality indices. This material forms the background of the later analysis.

I assume that society has preferences over alternative income distributions that can be represented by a social evaluation function (SEF), $W : \Omega \rightarrow \mathbb{R}$ where $\Omega$ will be a different subset of $\mathbb{R}^N$ depending on which type of index is used. $W$ is continuous, increasing along the ray of equality and has all social indifference curves crossing the ray
of equality at some point\(^1\). Social evaluation functions of this type are called regular. I can use the SEF to construct the equally distributed equivalent income function (EDE)\(^2\) defined implicitly by equation 2.1:

\[
\xi = \Xi(y) \leftrightarrow W(\xi 1^N) = W(y)
\]  

where \(1^N\) denotes the \(N\)-vector of ones. The EDE gives the per capita level of income which, if distributed equally throughout the population, would provide society with the same level of welfare as does the current income distribution. The EDE function is a continuous ordinal transform of the SEF. Thus, for two income distributions \(\hat{y}\) and \(\bar{y}\), \(\Xi(\hat{y}) \geq \Xi(\bar{y})\) if and only if \(W(\hat{y}) \geq W(\bar{y})\).

The EDE function is the foundation of the two types of ethical inequality indices that I consider in this paper. The Atkinson-Kolm-Sen (AKS) inequality index for an arbitrary SEF is

\[
I(y) = 1 - \frac{\Xi(y)}{\mu(y)}
\]  

where \(\mu(y)\) is the mean of the income vector \(y\) and \(\Omega = \{R^N \mid \sum y_i > 0\}\). The AKS inequality index \(I(y)\) has many well known specific inequality indices as special cases. Setting \(\Xi(y) = (\sum_{i=1}^N (2i - 1)\hat{y}_i)/N^2\), where \(\hat{y}\) is a ranked permutation of \(y\) such that \(\hat{y}_i \geq \hat{y}_{i+1}\), gives the Gini inequality index \(I^G(y)\), setting \(\Xi(y) = \mu(y) - \sum_{i=1}^N (y_i/N) \log(y_i/N)\) produces the Theil inequality index and setting \(\Xi(y) = \mu(y) - [\sum_{i=1}^N (y_i - \mu(y))^2/N]^{1/2}\) yields the coefficient of variation inequality index. Corresponding to the AKS inequality index is the AKS equality index, defined by \(E(y) = 1 - I(y) = \Xi(y)/\mu(y)\).

The AKS index is the percentage of aggregate income that can be given up by the society and still be able to obtain the same level of well-being as with the original distribution by distributing the remainder equally among the individuals in the society.

\(^1\)The ray of equality is the ray defined by the equation \(y = \alpha 1^N\), where \(1^N\) is an \(N\)-vector of ones.

\(^2\)See Sen (1973) or Atkinson (1970) for more on the equally distributed equivalent income.
Thus it can be interpreted as the percentage of total income that is wasted because of inequality and does not contribute to social welfare.

The AKS index is the most commonly used type of inequality index in applied work but there is another general inequality index. This index is the Kolm-Blackorby-Donaldson (KBD) inequality index defined by

\[ A(y) = \mu(y) - \Xi(y) \]  

(2.3)

where in this case \( \Omega = \{ y \mid \sum y_i > 0 \} \) The KBD index gives the amount of income per capita that could be given up from the current aggregate income, provided that the remainder is distributed equally, and still preserve the same level of societal well being. It is the dollar value of the loss of potential welfare arising from income inequality.³

An index with the feature that \( I(y) = I(\lambda y) \) for every positive constant \( \lambda \) is termed a relative index. A relative index is used if the society is concerned only about percentage differences in income, or income shares, and not the actual differences in income. For example a change in the units that income is measured in has no effect on inequality measured by a relative index. It has been shown by Blackorby, Donaldson and Auersperg (1981) that the AKS index in equation 2.2 is a relative index if and only if the SEF \( W \) is homothetic.

Similarly an index for which \( A(y) = A(y + \delta 1^N) \), where \( \delta \) is a scalar, and \( y + \delta 1^N \in \Omega \), is called an absolute index. An absolute index is appropriate for a society which is concerned only with the differences in income measured in specific units. It can also be shown that the KBD index in equation 2.3 is an absolute index if and only if \( W \) satisfies a condition called translatability.⁴ A function \( W \) is translatable if and only if the SEF satisfies the following condition, \( W(y) = \phi(\bar{W}(y)) \) where \( \bar{W}(y) \) satisfies \( \bar{W}(x + \delta 1^N) = \bar{W}(x) + \delta \) for

³Because the KBD index measures inequality in dollars, intertemporal comparisons require \( y \) to be a vector of real incomes.
⁴See Chakravarty (1990) and Blackorby and Donaldson (1980).
2.3 Previous Decomposition Attempts.

In this section I discuss previous attempts at decomposing income inequality measures and provide examples to demonstrate some undesirable features of these decomposition methods. In subsequent sections I develop an alternative decomposition technique that does not share these drawbacks.

Income received as payment for a particular service or as a return to a particular asset is defined to be the factor component of that service or asset in income. The service or asset is the source of that factor component. Assume that an economy contains $J$ sources of income. The $j$'th income component of the income distribution is a vector $y^j \in \mathbb{R}^N$, where $N$ is the number of individuals in the economy. The $i$'th element of $y^j$, denoted $y^j_i$ gives the income of person $i$ from income source $j$. Let the total income distribution in the economy be denoted by $y \in \mathbb{R}^N$ where $y = \sum_{j=1}^J y^j$.

The Gini index of inequality is the most familiar inequality index to most economists. The Gini index can be written as

$$I(y) = 1 - \frac{\sum_{i=1}^N (2i - 1)\tilde{y}_i}{N^2 \mu(y)}$$

where $\tilde{y}$ is a permutation of the income vector $y$ such that $\tilde{y}_i \geq \tilde{y}_{i+1}$, $i = 1, ..., N$ and $\mu(y)$ is the mean income. Fei, Ranis and Kuo (1978) decompose the Gini index in the following way,

$$I^G(y) = \sum_{j=1}^J \frac{\mu(y^j)}{\mu(y)} S^{PG}(y^j, y)$$

Where $S^{PG}(y^j, y) = [1 - \sum_{i=1}^N (2i - 1)\tilde{y}_i^j/\mu(y^j)N^2]$. The term in the square brackets is known as the pseudo-Gini for income source $j$. The pseudo-Gini is similar to the Gini but
the rank assigned to person i’s income from source j is person i’s rank in total income, not his or her rank in the distribution of factor component j. The value of the pseudo-Gini for factor source j, weighted by the percentage of total income coming from source j, is defined to be the amount of inequality due to source j.

To see the problem with this decomposition of the Gini index consider the following example with two people and two sources of income. Let the vector of incomes from source one be y^1 = [1, 2] and from source two be y^2 = [2, 1]. Thus total income is a vector y = [3, 3]. The decomposition of the Gini index using the technique of Fei, Ranis and Kuo gives \( \mu(y^1) = \mu(y^2) = .5 \), \( S^{PG}(y^1, y) = -1/6 \) and \( S^{PG}(y^2, y) = 1/6 \) or vice-versa. Notice that the value of the pseudo-Gini is negative for the first source and positive for the second source even though one is just a permutation of the other. The reason for this is the arbitrary way of breaking a tie in the ranking of total income. In this case, the signs of the pseudo-Gini’s for the individual factor component distributions can be reversed simply by reversing the method used to break the tie.

The problem above is related to the continuity of the pseudo-Gini. The pseudo-Gini is not continuous in the factor component at the amount of income that will generate a tie in the amount of total income between individuals. Consider the above distribution of income with the exception that y^2 becomes y^2 = [2, 1 + \( \epsilon \)]. Then as \( \epsilon \) approaches 0 from below, the pseudo-Gini for income source 2 approaches 1/6 and for income source 1 the pseudo-Gini approaches -1/6. As \( \epsilon \) approaches 0 from above the pseudo-Gini for income source 2 approaches -1/6 and for income source 2 the pseudo-Gini approaches 1/6.

Layard and Zabalza (1979) use the variance as a measure of income inequality. The following decomposition is obvious

\[
\text{Var}(y) = \text{Var}(y^1) + \text{Var}(y^2) + 2\text{Cov}(y^1, y^2) \tag{2.6}
\]
This decomposition has a drawback in that it is only valid when either the variance or coefficient of variation is the measure of inequality. A second problem is that, when using this decomposition, the number of terms in the decomposition increases with the number of income sources at a greater than proportional rate because the covariance between each possible pair of factor components must be included.

The next decomposition rule I examine in this section is the general decomposition rule proposed by Shorrocks (1982). In this paper Shorrocks states that the decomposition rules described above are not unique, then sets out six axioms that do lead to a unique decomposition rule. Of these six axioms there are two that, when combined, lead to a result that is undesirable. As well, a third axiom that Shorrocks uses casts doubt on the reasonableness of this decomposition for a broad class of inequality indices.

The six axioms used by Shorrocks are

**Assumption 1:** $I(y)$ is a continuous and symmetric inequality index.

**Assumption 2:**

a) $S_j(y^1, ..., y^J; J)$ is continuous in $y^j$,
b) $S_j(y^1, ..., y^J; J) = S_{\pi_j}(y^{\pi^1}, ..., y^{\pi^J}; J)$ if $\pi^1, ..., \pi^J$ is any permutation of $1, ..., j$.

**Assumption 3:** $S_1(y^1, ..., y^J; J) = S_1(y^1, y - y_1; 2) = S(y^1, y)$ and $S_j(y^1, ..., y^J; J) = S(y^j, y)$.

**Assumption 4:** $\sum_{j=1}^{J} S(y^j, y) = I(y)$

**Assumption 5:**

a) If $P$ is any $N \times N$ permutation matrix, then $S(y^1P, yP) = S(y^1, y)$ and b) $S(\nu^j 1^N, y) = 0$ for all $\nu^j$.

**Assumption 6:** $S(y^1, y^1 + y^1P) = S(y^1P, y^1 + y^1P)$ Given these assumption Shorrocks’ main result may be stated

**Proposition 1** Defining the contribution of source $j$ to inequality to be $S(y^j, y)$, then assumptions 1–6 imply

$$f_j = \frac{S(y^j, y)}{I(y)} = \frac{\text{Cov}(y^j, y)}{\text{Var}(y)}$$  \hspace{1cm} (2.7)
Shorrocks' result is that the percentage contribution to inequality of a given income source, \( f_j \), is equal to the value of the slope coefficient in the ordinary least squares regression between total income and the income source \( j \). Thus the percentage contribution of a given source is the same for all inequality indices. Assumption 4 is called Consistent Decomposition. It states that the sum of the terms in any decomposition must equal the value of the inequality index for total income. Assumption 6 is called Two Factor Symmetry. Now consider again the previous example. The overall inequality index equals zero and \( y^2 = y^1 \mathbf{P} \); therefore consistent decomposition and two factor symmetry lead to the conclusion that \( S(y^1, y) = S(y^2, y) = 0 \). This is so even though it is obvious that there is inequality present in the distribution of income from each source and without a given source there is inequality. Shorrock's axiom means that his decomposition ignores information such as this. It would be nice to have a decomposition rule which provided an indication of the amount of inequality present in the component distributions.

Consistent decomposition and two factor symmetry combine to produce an undesirable result but not an extremely bad result. The third axiom of Shorrock's that I consider in detail is difficult to reconcile with relative indices of inequality although it does seem to be reasonable for an absolute index of inequality. This axiom is assumption 5, part b and is called Normalization for Equal Factor Distributions. This axiom states that the contribution to inequality of an equally distributed factor component is zero. At first glance this may sound reasonable but consider what happens in the following example using the mean of order 1/2 index of relative inequality. Let \( y^1 = [1, 2] \), \( y^2 = [1, 1] \). Then \( S^S(y^1, y) = I(y) = .01 \). Now change \( y^2 \) to \( \hat{y}^2 = [2, 2] \). Again, by Normalization for Equal Factor Components, \( S^S(y^1, y) = I(y) \) but now note that \( S^S(y^1, y) = .005 \).
Out of two sources of income, only one has changed but this does not affect the contribution to inequality of the changed income source but does change the contribution to inequality of the factor component distribution that does not change. This is, perhaps, counterintuitive. The same objection does not apply to the Gini absolute index of inequality. Consider the previous example. In the first case $A(y) = A(y^1) = 1/4$ and in the second case, since the only change is that everyone's income has increased by the same amount, $A(y) = A(y^1) = 1/4$. Thus for the KBD absolute index this axiom seems more appropriate.

The last decomposition rule that I examine is for the Theil inequality index. The Theil Inequality index is given by the following equation

$$ I^T(y) = \sum_{i=1}^{N} \frac{y_i}{N\mu} \log \left( \frac{y_i}{\mu} \right) $$

West and Theil (1991) decompose equation 2.8 in the following way

$$ I^T(y) = \sum_{i=1}^{J} \frac{\mu^i}{\mu} \sum_{i=1}^{N} \frac{y_i^2}{N\mu^2} \log \left( \frac{y_i^2}{\mu^2} \right) - \sum_{i=1}^{J} \sum_{i=1}^{N} \frac{y_i^2}{N\mu} \log \left( \frac{y_i^2\mu}{Ny_i\mu^2} \right) $$

Notice that the first term is a weighted summation of Theil indices defined for the component incomes. The second term is an interaction term that gives a measure of how the various factor component distributions cancel out or reinforce the inequality present in the other distributions. Thus the philosophy of this decomposition is similar to that for the coefficient of variation inequality index.

### 2.4 Decomposing Inequality Indices

In this section I present two general methods for decomposing both the AKS and KBD inequality indices for arbitrary SEF's by income source. The method used is to first decompose the EDE income and then to use this to decompose the respective indices. In decomposing the EDE functions two basic methods are suggested, the first method
produces what I call an interactive decomposition and the second method produces what I call a Shapley decomposition.

2.4.1 Interactive Decompositions

In this subsection I discuss and derive a decomposition which includes a direct effect on inequality for each income source and an interaction effect similar to a covariance term in the standard decomposition of the variance. The approach taken is to state some properties that would be nice for a decomposition to have and then to propose some candidate decompositions. I make no claim that this characterizes all possible decompositions with these features but instead provides one intuitively appealing decomposition.

The initial feature I would like my decomposition to have is that it has numerical significance. That is I want a decomposition where it makes sense to speak of a particular element causing a particular percentage of inequality. The immediate effect of this requirement is that the decomposition must therefore be additive. Any other structure (such as a multiplicative one) means that the effect of one element of the decomposition cannot be separated from the effect of any of the other elements and therefore a given element cannot be said to contribute a certain percentage of inequality.\footnote{I have not claimed that additivity is \textit{sufficient} for numerical significance of the decomposition, only that it is \textit{necessary}.}

The second feature that I would like to have the decomposition satisfy is that the contribution of a given income source should depend, at least partially, on the inequality present in the income source itself, that is I would like the contribution of a source to inequality to be an increasing function of the amount of inequality present in the source. This seems to be a reasonable requirement for a decomposition.

Finally, consider what would happen if society starts initially with $J$ sources of income and eliminates inequality in each source of income in turn with the following thought...
experiment; regard each source of income as the only source of income and construct the EDE income for this source in the usual way. This gives a measure of the welfare that society would receive from income source $j$ if that source were the only possible source. This could then be considered a measure of the direct effect of income source $j$ on total welfare. It is extremely unlikely that the contributions of individual income sources will sum to the actual level of welfare because, in the aggregate distribution, some of the reduction in welfare because, for example, person $i$ has a very low amount of income source $j$ may be counteracted because person $i$ has a very high amount of income from source $j'$. This effect is not accounted for in summing the direct effect on welfare over all income sources so, besides the direct effect on welfare, there will be an interaction effect as well which takes this into account. It seems reasonable that I ask this decomposition of inequality indices to include a term similar to this interaction effect as well.

I now provide a candidate decomposition of the AKS and KBD inequality indices reflecting the three requirements above. Later in this subsection I discuss whether this is a reasonable decomposition in light of the suggested requirements.

Recall that the EDE income for a given income distribution is given by the function $\xi = \Xi(y)$. The EDE function can be trivially rewritten as

$$\Xi(y) = \sum_{j=1}^{J} \Xi(y^j) + \Xi(y) - \sum_{j=1}^{J} \Xi(y^j),$$

which, defining $\mu^j = \mu(y^j)$, $\mu = \mu(y)$ and $s^j = \mu^j / \mu$, where $s^j$ is the share of source $j$ in total income, can in turn be rewritten as

$$\Xi(y) = \sum_{j=1}^{J} \Xi(y^j) + \sum_{j=1}^{J} [s^j \Xi(y) - \Xi(y^j)].$$  

This decomposition of $\Xi$ reflects requirement three above. To construct the decomposition of the AKS inequality index, multiply equation 2.11 by $-1/\mu$ and add 1 to both
sides to obtain

\[ I(y) = 1 - \sum_{j=1}^{J} \frac{\Xi(y^j)}{\mu} - \sum_{j=1}^{J} \left[ s^j \frac{\Xi(y)}{\mu} - \frac{\Xi(y^j)}{\mu} \right], \quad (2.12) \]

which can be written as

\[ I(y) = 1 - \sum_{j=1}^{J} s^j \frac{\Xi(y^j)}{\mu^j} - \sum_{j=1}^{J} \left[ s^j \frac{\Xi(y)}{\mu} - s^j \frac{\Xi(y^j)}{\mu^j} \right]. \quad (2.13) \]

Remembering that \( \sum_{j=1}^{J} s^j = 1, \)

\[ I(y) = \sum_{j=1}^{J} s^j I(y^j) + \sum_{j=1}^{J} s^j [I(y) - I(y^j)]. \quad (2.14) \]

Now, defining \( C^I(y^1, \ldots, y^J, y) = \sum_{j=1}^{J} [I(y) - I(y^j)] \), equation 2.14 may be rewritten more compactly as

\[ I(y) = \sum_{j=1}^{J} s^j I(y^j) + C^I(y^1, \ldots, y^J, y). \quad (2.15) \]

The first term in equation 2.15 is a weighted sum of the relevant inequality indices for the different factor component distributions, with weights equal to the fraction of total income generated by source \( j \). Note that these are the actual indices, not pseudo-indices as in Fei, Ranis and Kuo (1978) and thus do not depend on total income \( y \). The term \( C^I(y^1, \ldots, y^J, y) \) is an interaction term which measures how much inequality in the \( J \) factor component distributions is counteracted by inequality in the other \( J - 1 \) distributions of factor component incomes. If this value is negative then it means that the interaction is income equalizing and if it is positive then the interaction is income disequalizing.

The KBD inequality indices can be decomposed in a similar way to the AKS. Recall from equation 2.11

\[ \Xi(y) = \sum_{j=1}^{J} \Xi(y^j) + \sum_{j=1}^{J} [s^j \Xi(y) - \Xi(y^j)], \quad (2.16) \]

the KBD index is defined as \( A(y) = \mu - \Xi(y) \) so substituting in from 2.16 I get

\[ A(y) = \mu - \sum_{j=1}^{J} \Xi(y^j) - \sum_{j=1}^{J} [\mu - \mu + s^j \Xi(y) - \Xi(y^j)], \quad (2.17) \]
which can be rewritten as

\[ A(y) = \sum_{j=1}^{J} \mu^j - \Omega(y^j) \]  

and now substituting in from the definition of the KBD index leaves the decomposition

\[ A(y) = \sum_{j=1}^{J} A(y^j) + \sum_{j=1}^{J} s^j (A(y) - A(y^j)). \]  

The decomposition in equation 2.19 is again a weighted sum of the relevant inequality indices for the different distributions of factor component income plus an interaction term. In contrast to the AKS index, however, the weights on the inequality indices in the KBD decomposition are equal to one. This is intuitively because the units that the indices \( A(y) \) and \( A(y^j) \) are measured in are the same, whereas in the case of the AKS index, the units of \( I(y) \) and \( I(y^j) \) are different. \( I(y) \) is measured in terms of the fraction of total income but \( I(y^j) \) is measured in terms of the fraction of income from source \( j \). Both \( A(y) \) and \( A(y^j) \) are measured in terms of income.

Defining \( C^A(y^1, ..., y^J, y) = \sum_{j=1}^{J} [s^j A(y) - A(y^j)] \), equation 2.19 can be expressed more compactly as

\[ A(y) = \sum_{j=1}^{J} A(y^j) + C^A(y^1, ..., y^J, y). \]  

The term \( C^A(y^1, ..., y^J, y) \) is again an interaction term composed of a scaled down KBD index for total income and a KBD index for income from source \( j \). The scaling down of \( A(y) \) is done by multiplying by the share of total income arising from source \( j \). This apportions the EDE income in a way consistent with every factor component distribution having the same pattern of inequality. If \( C^A(y^1, ..., y^J, y) \) is positive, this has the interpretation that the interaction between income sources reinforce the inequality present in the distribution of income.

The previous decompositions of the AKS and the KBD inequality indices are based on considering the individual factor component distributions. Suppose the interesting
question is not how much a given source of income contributes to inequality but instead is how does inequality change as a result of income source \( j \). The current decompositions do not necessarily answer this question. An alternative, and perhaps equally plausible, decomposition can be obtained by using the distribution of income from all sources except source \( j \). Denote by \( y^{-j} \) the distribution of income from all sources but source \( j \), that is \( y^{-j} = y - y^j \).

I follow the same general method for decomposing the EDE income. Thus the counterpart to equation 2.11 is

\[
\Xi(y) = \sum_{j=1}^{J} \Xi(y^{-j}) - \sum_{j=1}^{J} \Xi(y^{-j}) + \Xi(y) .
\]  

(2.21)

Given 2.21 the AKS inequality index is

\[
I(y) = 1 - \sum_{j=1}^{J} \frac{\mu^{-j} \Xi(y^{-j})}{\mu^{-j}} - \sum_{j=1}^{J} \frac{\mu^{-j} \Xi(y^{-j})}{\mu^{-j}} + \frac{\Xi(y)}{\mu} .
\]  

(2.22)

Multiplying terms by \( \mu^{-j} / \mu^{-j} \) yields

\[
I(y) = 1 - \sum_{j=1}^{J} s^{-j} \frac{\Xi(y^{-j})}{\mu^{-j}} + \sum_{j=1}^{J} s^{-j} \Xi(y^{-j}) - \frac{\Xi(y)}{\mu} .
\]  

(2.23)

or

\[
I(y) = 1 - \sum_{j=1}^{J} (1 - s^j) \frac{\Xi(y^{-j})}{\mu^{-j}} + \sum_{j=1}^{J} (1 - s^j) \Xi(y^{-j}) - \frac{\Xi(y)}{\mu} ,
\]  

(2.24)

Now by definition \( s^{-j} = 1 - s^j \) so by substitution

\[
I(y) = 1 - \sum_{j=1}^{J} (1 - s^j) \Xi(y^{-j}) + \sum_{j=1}^{J} (1 - s^j) \Xi(y^{-j}) - \frac{\Xi(y)}{\mu} ,
\]  

(2.25)

or by adding and subtracting one and rearranging

\[
I(y) = \sum_{j=1}^{J} s^j I(y^{-j}) + \sum_{j=1}^{J} s^j [I(y) - I(y^{-j})] .
\]  

(2.26)

Which is more compactly expressed as

\[
I(y) = C_I(y^{-1}, ..., y^{-J}, y) + \sum_{j=1}^{J} s^j [I(y) - I(y^{-j})] .
\]  

(2.27)
Equation 2.27 is the final decomposition of the AKS index $I(y)$. The terms in this decomposition have the following interpretation. The first term is a weighted sum of the relevant inequality index defined over the distributions $y^{-j}$. This provides a base level of inequality and is similar to the interaction term from the previous direct interactive decomposition. The second term is the difference between the inequality in the distribution $y$ and the distribution $y^{-j}$ and is the contribution to total inequality of source $j$ given the distribution $y^{-j}$. It thus represents the *marginal* contribution of source $j$ to inequality. If the inclusion of the $j$th factor component results in a more unequal distribution then this second term is positive and $I(y)$ is increased by $y^j$. The value of this marginal term is positive if income source $j$ is unequalizing and is negative if income source $j$ is equalizing.

The KBD index can be decomposed in a similar way. Again starting with the decomposition of the EDE income

$$\Xi(y) = \sum_{j=1}^{J} \Xi(y^{-j}) - \sum_{j=1}^{J} \Xi(y^{-j}) + \Xi(y)$$  \hspace{1cm} (2.28)

Adding and subtracting $\sum_{j=1}^{J} \mu^{-j}$ gives the expression

$$\Xi(y) = \sum_{j=1}^{J} [\mu^{-j} - \Xi(y^{-j})] - \sum_{j=1}^{J} [\mu^{-j} - \Xi(y^{-j})] + \Xi(y)$$  \hspace{1cm} (2.29)

Then recalling the definition of the KBD index gives

$$A(y) = \sum_{j=1}^{J} A(y^{-j}) - \sum_{j=1}^{J} A(y^{-j}) + A(y)$$  \hspace{1cm} (2.30)

Rewriting equation 2.30 leaves

$$A(y) = \sum_{j=1}^{J} A(y^{-j}) + \sum_{j=1}^{J} [s^j A(y) - A(y^{-j})]$$  \hspace{1cm} (2.31)

or

$$A(y) = C^A(y^{-1},...,y^{-J},y) + \sum_{j=1}^{J} [s^j A(y) - A(y^{-j})]$$  \hspace{1cm} (2.32)
Equation 2.32 has an interpretation similar to the interpretation of equation 2.27. The first term is a base level of inequality generated by the distributions. This base level is then modified by the second term which again has the interpretation as the marginal effect of the jth source of income on the inequality distribution. If this value is positive, then the marginal effect of introducing the jth source of income is to increase the inequality in the total distribution.

To help in justifying the decomposition in this section consider the generalization of the decomposition 2.15 for the relative AKS inequality index given by

\[ I(y) = \sum_{j=1}^{J} a^j I(y^j) + \sum_{j=1}^{J} b^j I(y) - \sum_{j=1}^{J} a^j I(y^j), \]  

(2.33)

where \( \sum_{j=1}^{J} a^j = \sum_{j=1}^{J} b^j = 1 \). Now take the example with two individuals and J income sources. The vector of incomes from source 1 is \( y^1 = [y_1^1, y_2^1] \), from source j, \( j \neq 1 \) is \( y^j = [\alpha^j y_1^1, \alpha^j y_2^1] \), and the vector of total income is

\[ y = [(1 + \sum_{j=2}^{J} \alpha^j)y_1^1, (1 + \sum_{j=2}^{J} \alpha^j)y_2^1]. \]  

(2.34)

Thus \( I(y^1) = I(y^2) = I(y^1) \) and the interaction term in the decomposition is identically zero. What then would be the percentage contribution to total inequality assigned to income source j? It would be equal to \( a^j I(y^j)/I(y) = a^j \). Now since the measured inequality in each source \( y^j \) is the same, it seems reasonable that the percentage contribution to inequality be equal to the percentage of total income supplied by income source j. For this to be true for all vectors of income it must be true that \( a^k = \alpha_k / (1 + \sum_{j=2}^{J} \alpha^j) = s^i \) for \( k \neq 1 \) and \( a^1 = 1/(1 + \sum_{j=2}^{J} \alpha^j) = s^i \). The value of the interaction term \( C(y^1, ..., y^J) \) is unchanged with different values of \( b^j \), and so there is no harm in setting \( b^j = s^i \) as well.

The AKS inequality index is a cardinal measure of inequality, that is it makes sense to say that one distribution is twice as unequal as another distribution. Does this imply
anything about the decomposition? Cardinality means that the two functions $I(y)$ and $\lambda I(y)$, $\lambda > 0$ are informationally equivalent. Thus the decompositions

$$I(y) = \sum_{j=1}^{J} a^j I(y^j) + \sum_{j=1}^{J} b^j I(y) - \sum_{j=1}^{J} a^j I(y^j), \quad (2.35)$$

and

$$\lambda I(y) = \sum_{j=1}^{J} \lambda a^j I(y^j) + \sum_{j=1}^{J} \lambda b^j I(y) - \sum_{j=1}^{J} \lambda a^j I(y^j) \quad (2.36)$$

are informationally equivalent and the individual terms of the decomposition have numerical meaning. The implication is that it makes sense to speak of source $j$ contributing a given percentage to inequality.

I now present an example of the decompositions which should help highlight the intuition behind the interactive decompositions. I will focus on the AKS index for the SEF, mean of order one half\(^7\). Consider a society of two individuals who each have two possible sources of income. The first source is the return to personal characteristics which are not education, such as experience. Suppose individual 1 has a return to experience of $\$2$ and the second person has a return to experience of $\$4$. Denote this income source 1 and it is given by the vector $y^1$, $y^1 = [2, \ 4]$. The second source of income is the returns to education. Suppose individual 1 has a return to education of $\$1$ and individual two has no income from education so the vector of income from source 2 is $y^2 = [1, \ 0]$. Overall income is the sum of these two vectors so is $y = [3, \ 4]$.

The above vectors give values for $\Xi(y) = 3.4821$, $\Xi(y^1) = 2.9142$, and $\Xi(y^2) = .25$. The direct interactive decomposition yields a contribution of .0245 for source 1, .0714 for source 2, and a interaction term of .0908. The intuition behind these results are

\(^7\)The mean of order 1/2 index is derived from the following EDE function

$$\xi = \left[ \frac{y_{1}^{1/2} + y_{2}^{1/2}}{2} \right]^2.$$
that source two contributes more to total inequality than source 1 does. This is so, even though source 2 is a much less important source of income than source 1 (in terms of magnitude), because the measured inequality in source 2 is much higher than that in source 1. The inequality in source 1 counteracts the inequality of income from source 2, and vice versa. Overall the interaction of the two sources reduces measured inequality by .0908.

The marginal interactive decomposition gives a value of -.4242 for the marginal contribution of source 1, -.0034 for the marginal contribution of the returns to education, and a base term of .4327. Thus here we see that source 1 is also a much more important source of reductions in inequality than is source 2. This results from a combination of the addition of source 1 counteracting some of the inequality in source 2, and the dramatic increase in mean income when we add source 1. The increase in mean income means that for the same difference in levels of income, the measured inequality is lower. The cross term of .4327 provides a base from which the marginal effects are measured.

The previous discussion highlights the reasons why the interactive decomposition takes the form that it does instead of some other similar form. Although I concentrated on the interactive decomposition given in equation 2.15, the same procedure will produce similar justifications for the other interactive decompositions.

2.4.2 Shapley Decomposition

The interactive decompositions presented in the previous subsection are intuitively appealing and have some well known decompositions as special cases. In one way however they may not provide us with the answer that we want. They provide only a partial answer to the question of how much of inequality is a result of a given income source. This problem arises because of the non-separability of the interaction effect. I now present a complete characterization, based on the Shapley value of transferable utility games, of
an ethical decomposition that does allow for unambiguous allocation of the effects.  

I again assume $J$ income sources in random order $j = 1, \ldots, J$. Note vectors of income do not have to be independent of each other, only that they be in random order. The vector of incomes from source $j$ is the $N$-vector $y^j$. The aggregate income distribution is defined by $y = \sum_{j=1}^{J} y^j$.

Let $M$ denote the set of vectors of income sources, that is $M = \{y^1, y^2, \ldots, y^J\}$. A subset of income sources is denoted $C \subset \{1, \ldots, J\}$. The grand subset containing all elements of $\{1, \ldots, J\}$ will be denoted $C^J$. Denote by $c = |C|$ the number of elements in the subset $C$. For a given subset the aggregate income from subset $C$ is denoted $y^C = \sum_{j \in C} y^j$.

In an ethical decomposition of an inequality index the first step is to decompose the EDE income function $\Xi(y)$. Let $\sigma_C(M, \Xi)$ be the contribution of subset $C$ to the EDE income determined by the function $\Xi$ and the set $M$. Denote by $\sigma_j$ the contribution of source $j$ to the value of the EDE. Let the sum all $j$ of the contributions equal the EDE income $\Xi(y)$. Thus

$$\sum_{j=1}^{J} \sigma_j(M, \Xi) = \Xi(y) \quad (2.37)$$

Since the SEF and thus the EDE income function are functions of aggregate income only, the decomposition of the EDE function should be symmetric with respect to income sources. This means that the name of the income source or how it was earned does not matter. Formally, symmetry with respect to income sources requires

**Axiom A)** For any two sets of income sources, $M$ and $M^\pi$, where $M^\pi$ is a permutation
of $M$, and income source $k$ in $M$ is $\pi k$ in $M^\pi$

$$\sigma_k(M, \Xi) = \sigma_{\pi k}(M^\pi, \Xi) \quad (2.38)$$

The second axiom is borrowed from Shorrocks (1982) who calls it independence of the level of disaggregation. Suppose there are two income sources from the same population. This axiom states that the sum of the individual contributions must equal the total contribution of a new income source constructed by adding the original two sources together. Formally

**Axiom B** For any two subsets of income sources $C'$ and $C''$

$$\sigma_C(M, \Xi) = \sigma_{C'}(M, \Xi) + \sigma_{C''}(M, \Xi) \quad (2.39)$$

where $C = C' \cup C''$, and $C' \cap C'' = \emptyset$.

Define the welfare from a subset $y^C$ to be equal to $\Xi(y^C)$ Now define an income source's marginal contribution to subset $C$ as $\Xi_j(y^C) = \Xi(y^C) - \Xi(y^C - y^j)$. The marginal contribution is a measure of the change in the welfare from a particular subset, that is produced when the $j$th income source is included compared to when it is excluded. When the marginal contribution is positive then the income source would increase the welfare from the subset and when it is negative then the source reduces welfare from the given subset. In general the counteraction of inequality in one source by inequality in other sources will mean that the marginal contribution of a source is not equal to the welfare from the source.

The last axiom that I use in deriving the Shapley decomposition is that the contribution of source $j$ to inequality be a function of the marginal contributions of source $j$ to the various subsets of income sources. What is ultimately of concern is the total effect that an income source has on welfare. The simplest measure of this effect is comparing welfare with, and without, the given income source\textsuperscript{11}. The most fundamental measure

\textsuperscript{11}Such as in the Marginal interactive decomposition.
of this effect is the marginal contribution, to total welfare from all sources of income, of a given income source. But that is not the only way that a source can contribute to measured welfare. Each income source interacts with all other income sources both individually and in combination to produce the total measure of welfare. For example returns to education can counteract inequality in transfer payment income by itself, can reinforce inequality in non-labour income by itself, and could have a neutral effect on the inequality in transfer payment plus non-labour income. All three effects are legitimate components of how returns to education contribute to overall welfare. Thus a decomposition to find the total effect of a given income source on welfare should consider the marginal effect of a given income source on all possible subsets of income sources. Axiom C puts a little more structure on the way in which the marginal effect contributes to the decomposition.

**Axiom C** For two sets of income sources $M$ and $\bar{M}$, with $y^C$ the vector of incomes from coalition $C$ in $M$, and $\bar{y}^C$ the vector of incomes from coalition $C$ in $\bar{M}$, if $\Xi^j(y^C) \geq \Xi^j(\bar{y}^C)$ for all $C$ then $\sigma_j(M, \Xi) \geq \sigma_j(\bar{M}, \Xi)$; or

b) for a given set of income sources and two EDE functions $\Xi_1$ and $\Xi_2$, if $\Xi_1^j(y^C) \geq \Xi_2^j(y^C)$, for all $C$ then $\sigma_j(M, \Xi_1) \geq \sigma_j(M, \Xi_2)$.

The first part of this axiom says that if there are two different economies with different sets of income sources, and if an income source consistently contributes a larger marginal effect to all subsets from one economy than to subsets from the other economy, then the contribution of that income source to welfare in the first economy must be greater than it’s contribution in the second economy. The second part of this axiom may also need some explanation. As a society, it is by no means obvious what our social preferences are. This part of the axiom says that if we compare the same economy with two different social preferences and one set of preferences consistently attributes a higher marginal effect to
Chapter 2. Decomposing Inequality Indices by Income Source

one income source, then the decomposition of welfare using the social preferences which attributes a higher marginal effect will also yield a higher contribution to total welfare for the given income source.

The next step is to construct an equivalent function to \( \Xi(y^C) \), denoted \( \hat{\Xi} \), which is defined on the power set of \( C^J \) and the set of income sources \( M \), for which \( \hat{\Xi}(C, M) = \Xi(y^C) \) for all \( C \). Consider the function

\[
\bar{\Xi}(C, M) = \Xi(I_1^Cy_1^1 + I_2^Cy_2^2 + \ldots + I_J^Cy_J^J)
\]  

(2.40)

where

\[
I_j^C = \begin{cases} 
1 & \text{if } j \in C, \\
0 & \text{otherwise}
\end{cases}
\]

(2.41)

It is easy to see that \( \bar{\Xi}(C, M) = \Xi(y^C) \) for all \( C \). Thus any decomposition of \( \hat{\Xi} \) will imply an equivalent decomposition of \( \Xi \). I will therefore concentrate on a decomposition \( \hat{\sigma}(C^J, \hat{\Xi}, M) \) of \( \hat{\Xi} \). Now consistency requires the sum of the \( \hat{\sigma}_j \) must equal \( \hat{\Xi} \).

\[
\sum_{j=1}^{J} \hat{\sigma}_j(C^J, \hat{\Xi}, M) = \hat{\Xi}(C^J, M)
\]

(2.42)

In turn the axioms A, B, and C have the following implications for \( \hat{\sigma}_j(\hat{\Xi}) \)

**Axiom A')** For any two vectors of income sources \( M \) and \( M^\pi \) where income source \( k \) in \( M \) is \( \pi k \) in \( M^\pi \), and \( C^{\pi J} \) is the appropriate permutation of the grand coalition,

\[
\hat{\sigma}_k(C^J, \hat{\Xi}, M) = \hat{\sigma}_{\pi k}(C^{\pi J}, \hat{\Xi}, M^\pi)
\]

(2.43)

**Axiom B')** \( \hat{\sigma}_C(C^J, \hat{\Xi}, M) = \hat{\sigma}_{C^\prime}(C^J, \hat{\Xi}, M) + \hat{\sigma}_{C^\prime^\prime}(C^J, \hat{\Xi}, M) \) where \( C = C^\prime \cup C^\prime^\prime \) and \( C^\prime \cap C^\prime^\prime = \emptyset \)

**Axiom C')

a) For any two sets of income sources \( M \) and \( \bar{M} \), \( \hat{\Xi}^j(C, M) \geq \hat{\Xi}^j(C, \bar{M}) \), for all \( C \) where \( \hat{\Xi}^j(C, M) = \hat{\Xi}(C, M) - \hat{\Xi}(C \setminus j, M) \) implies \( \hat{\sigma}_j(C^J, \hat{\Xi}, M) \geq \hat{\sigma}_j(C^J, \hat{\Xi}, M) \) and,
Chapter 2. Decomposing Inequality Indices by Income Source

b) For any given $M$ and two EDE functions $\tilde{\Xi}_1$ and $\tilde{\Xi}_2$, $\tilde{\Xi}_1(C, M) \geq \tilde{\Xi}_2(C, M)$ for all $C$ implies $\tilde{\sigma}_j(C^j_1, \tilde{\Xi}_1, M) \geq \tilde{\sigma}_j(C^j_2, \tilde{\Xi}_2, M)$.

Actually axiom $B'$ turns out to be redundant. Axioms $A'$ and $C'$ are enough to uniquely determine the decomposition rule as the Shapley rule. Since the Shapley rule does satisfy $B'$ there is no contradiction.

**Proposition 2** There is only one decomposition rule satisfying $A'$ and $C'$. It is the Shapley decomposition given by

$$\tilde{\sigma}_j = \sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \sum_{\{C \subseteq c, \forall \in C\}} \tilde{\Xi}(C)$$

(2.44)

**Proof:** The proof of the proposition is from Young (1985)\(^{12}\). QED

Since $\tilde{\Xi}(C, M) = \Xi(y^C)$ for all $C$, it must be the case that $\tilde{\sigma}_j(C^j, \tilde{\Xi}, M)$ is a decomposition of $\Xi(y)$ and it will satisfy the two axioms $A$, and $C$. It is easy to demonstrate that $\tilde{\sigma}_j(C^j, \tilde{\Xi}, M) = \sigma_j(M, \Xi)$ where

$$\sigma_j(M, \Xi) = \sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \sum_{\{C \subseteq c, \forall \in C\}} [\Xi(y^C) - \Xi(y^C - y^j)]$$

(2.45)

An alternative characterization of the Shapley value $\tilde{\sigma}_j$ given in Moulin (1988) demonstrates that $\tilde{\sigma}_j$ satisfies axiom $B'$ and thus $\sigma_j$ satisfies $b$.

A potential problem with the Shapley decomposition of welfare may be that not all subsets of income sources are feasible. For example consider the case where there are three sources of income which are returns to high school education, returns to post-secondary education, and returns to other potential factors\(^ {13}\). In this example it is unreasonable to

---

\(^{12}\)Actually Young (1985) assumes a stronger version of axiom $C'$ than I use but his proof only requires the weaker version.

\(^{13}\)This case is examined in the next chapter.
talk about positive returns to post-secondary education if the individual does not first have a return to high school education. This is admittedly a drawback in the analysis. In the case where not all subsets of income are reasonable it may be the case that the Shapley decomposition is not the only decomposition that satisfies the three axioms presented but it will still be the case that the Shapley decomposition does satisfy the axioms.

The decomposition of the EDE function allows a corresponding decomposition of the AKS and the KBD inequality indices. Recall that the AKS index is given by $I(y) = 1 - \frac{\Xi(y)}{\mu}$. Substitute the Shapley decomposition $\sigma_j$ into the AKS index to get

$$I(y) = 1 - \frac{\sum_{j=1}^{J} \sigma_j}{\mu}$$

(2.46)

Multiplying and dividing by $\mu^j$ gives

$$I(y) = 1 - \sum_{j=1}^{J} \frac{\mu^j \sigma_j}{\mu^j}$$

(2.47)

which in turn equals

$$I(y) = \sum_{j=1}^{J} \frac{\mu^j}{\mu} \left[ 1 - \frac{\sigma_j}{\mu^j} \right]$$

(2.48)

The formula in equation 2.48 can be broken down into two parts. The first part is the share of income source $j$ in total income. This is multiplied by the second part which resembles an AKS inequality index but instead of using the EDE function it uses the Shapley contribution of source $j$ to the EDE income.

Now I use the Shapley decomposition of the EDE function to decompose the KBD inequality index. Recall that the KBD index is given by $A(y) = \mu - \Xi(y)$ which can be written as

$$A(y) = \mu - \sum_{j=1}^{J} \sigma_j$$

(2.49)

or in turn

$$A(y) = \sum_{j=1}^{J} [\mu^j - \sigma_j]$$

(2.50)
Chapter 2. Decomposing Inequality Indices by Income Source

Equation 2.50 has the following interpretation. It is the sum of KBD inequality indices for individual income vectors with the modification that only the Shapley contribution of income source $j$ to the overall EDE income is used instead of the actual EDE income.

The intuition behind the Shapley decomposition can best be thought of in reference to the direct interactive decomposition. In the interactive decomposition the interaction term is present and makes it impossible to allocate all of the inequality to one source or another. The Shapley decomposition gets around this problem and one way to think of what it is doing is separating the interaction effect and allocating it to the individual components. To separate the effect of a given income source on the interaction effect, it must consider all possible sources of interaction. This is the reason for examining all possible coalitions of income sources.

The intuition for the complete decomposition provides some additional intuition about axiom C. What this axiom says is that if two economies are compared, and in one world a given source has a greater marginal effect on all sources of interaction than in the other world, then the total effect of that source in the first world must be greater than that in the second world.

The usefulness of the Shapley decomposition arises from two things. First many people may be uncomfortable using either of the interactive decompositions because of the presence of the interaction term which cannot be allocated between the different sources. The solution to such philosophical problems is to use a method which has no troublesome interactions. Although alternative methods to the Shapley decomposition exist the pseudo-gini approach has serious technical problems, and the Shorrocks decomposition has some properties that may make it undesirable for the AKS inequality indices. The Shapley decomposition is an attempt at providing a decomposition which avoids the problems of having an unpleasant interaction term but at the same time, since it is based on an explicit decomposition of the EDE function, should be equally applicable to AKS
As a help in understanding the Shapley decomposition I present the example with two individuals and two income sources, returns to education and returns to other sources, that I constructed previously. Recall that the vector of incomes from other personal factors, which I denote \( y^1 \) is \( y^1 = [2, 4] \), and the vector of income from education is denoted \( y^2 \), \( y^2 = [1, 0] \). The vector of total income is again \( y^1 = [3, 4] \). I again use the mean of order 1/2 SEF. With this SEF the EDE values are \( \Xi(\mathbf{y}) = 3.4821 \), \( \Xi(y^1) = 2.9142 \), and \( \Xi(y^2) = .2500 \). The Shapley contributions of the two sources to welfare are respectively \( \sigma_1 = 3.0732 \) and \( \sigma_2 = .4090 \). These numbers correspond to a contribution of source 1 to the AKS index of -.0290 and a contribution of returns to education to inequality of .0260, for a total AKS index of .0051. The interpretation of these contributions is that the returns to characteristics other than education reduce inequality. The intuition behind why this is true can be seen by examining the vectors of income sources themselves. Notice that while the absolute difference in income is greater for the first vector, the difference as a proportion of mean income from the source is greater for the second vector. The inequality index is a function both of the contribution to welfare from differences in income and the contribution to welfare from the mean of income. For income source 1 the increase in welfare because of the increased mean income outweighs the decrease in welfare from the greater differences in income. This yields an overall negative effect in the contribution of source 1 to lost welfare. Likewise the lower mean income outweighs the decreased income differences in the first source, leaving a positive contribution of this source to welfare loss.

The approach that I use in this paper is to first decompose the EDE income function and then use this decomposition to construct a decomposition of the inequality index. An alternative approach is to use the same axioms and to apply them directly to the inequality indices themselves. This will provide a contribution to the AKS index for the
first of two income sources of

\[
\sigma_j^I = \frac{1}{2} I(y^1) + \frac{1}{2} [I(y) - I(y^2)]
\]  

(2.51)

and of the KBD index

\[
\sigma_j^A = \frac{1}{2} A(y^1) + \frac{1}{2} [A(y) - A(y^2)].
\]  

(2.52)

Comparing this to the ethical Shapley decompositions

\[
s^I \left[ 1 - \frac{\sigma_j}{\mu^I} \right]
\]

(2.53)

and

\[
\mu^j - \sigma_j
\]

(2.54)

it is obvious that, in general, the results of the direct decompositions are not the same as the results of the Shapley decompositions presented in this paper. This is in contrast to the results of the interactive decomposition where a straightforward decomposition of the actual inequality indices can produce the same decomposition as the one that I obtain from the ethical approach.

### 2.5 A Comparison of the Decompositions.

In this section I consider the main features of the decompositions introduced in the previous sections. First the two general approaches, interactive and Shapley, are compared, then these two new approaches are compared with the previously known decomposition methods. I outline the main issues involved in choosing which of the decompositions is most appropriate in a given situation.

#### 2.5.1 Interactive versus Shapley Decompositions

The comparison of the two new decompositions suggested in this paper is easiest if I consider only two income sources. Because the Shapley decomposition of the EDE
income requires the calculation, for each income source, of the marginal contribution of that source to each possible coalition, it can require a lot of computational effort. The interactive decomposition always has $2J$ terms so the calculation of the contributions is in general much easier for the interactive decomposition than for the Shapley decomposition.

An index that is homogenous of degree zero in incomes is said to be a relative index. A desirable feature for any decomposition of a relative index is that each component of the decomposition is homogenous of degree 0 as well. Recalling the two decompositions of the AKS index

$$I(y) = \sum_{j=1}^{J} s^j I(y^j) - \sum_{j=1}^{J} s^j [I(y) - I(y^j)]$$

(2.55)

and

$$I(y) = \sum_{j=1}^{J} s^j I(y^{-j}) + \sum_{j=1}^{J} s^j [I(y) - I(y^{-j})]$$

(2.56)

If $I(y)$ is a relative index then so are $I(y^j)$ and $I(y^{-j})$. If each $y^j$ is multiplied by a constant $\lambda$ then the overall distribution is also multiplied by $\lambda$. Thus the numerator and the denominator of the shares $s^j$ are multiplied by $\lambda$ and the shares are homogenous of degree zero. Therefore the elements of the decompositions of the AKS index are homogenous of degree zero if and only if the overall index is homogenous of degree zero.

The Shapley decomposition is also a relative decomposition if and only if the AKS index is a relative index. To see this note that Blackorby, Donaldson, and Auersperg (1981) have shown that the AKS index is a relative index if and only if the EDE function is homogenous of degree one in income. If the EDE income itself is homogenous then the Shapley decomposition of the EDE income is also homogenous of degree one, because it is additive in homogenous functions. The mean of an income source is homogenous of degree one so the function $1 - \sigma_j/\mu^j$ is homogenous of degree zero. I showed above that the share of income source $j$ in total income is homogenous of degree zero. This shows

---

14See Blackorby and Donaldson (1978) or Chakravarty (1990).
that the Shapley decomposition of an AKS index is a relative decomposition if and only if the index itself is a relative index.

I turn now to the decomposition of the KBD index that I introduced in the previous section. Since the KBD index is, in general, used in different circumstances than the AKS index, it is natural that the decompositions should have different properties as well.

Recall that the decomposition of the KBD index can be written as

\[ A(y) = \sum_{j=1}^{J} A(y^j) + \sum_{j=1}^{J} [s_j^j A(y) - A(y^j)] \] (2.57)

and

\[ A(y) = \sum_{j=1}^{J} A(y^{-j}) + \sum_{j=1}^{J} [s_j^j A(y) - A(y^{-j})] \] (2.58)

Both of these are obviously continuous functions of both \( y^j \) and \( y \). As well it can quickly be verified that both decompositions are symmetric with respect to the different income sources.

Consider what happens if, given a distribution \( y \) with component distributions \( y^1 \) and \( y^2 \), a new total distribution \( \hat{y} \) is constructed so that \( \hat{y} = y + \theta^1 N \). Denote by \( \theta^j \) the amount of extra income attributed to source \( j \). If \( w \) is translatable then \( A(y) \) is an absolute index and \( A(\hat{y}) = A(y) \) and \( A(\hat{y}^j) = A(y^j) \). The first decomposition of \( A(\hat{y}) \) can be written as

\[ A(\hat{y}) = \sum_{j=1}^{J} A(\hat{y}^j) + \sum_{j=1}^{J} [s_j^j A(\hat{y}) - A(\hat{y}^j)], \] (2.59)

which equals

\[ A(\hat{y}) = \sum_{j=1}^{J} A(\hat{y}^j) + [A(\hat{y}) - \sum_{j=1}^{J} A(\hat{y}^j)]. \] (2.60)

Similarly for the second decomposition

\[ A(\hat{y}) = \sum_{j=1}^{J} A(\hat{y}^{-j}) + [A(\hat{y}) - \sum_{j=1}^{J} A(\hat{y}^{-j})]. \] (2.61)

Thus the decomposition of the KBD index is invariant with respect to the addition of the vector \([ \theta^1 \; \ldots \; \theta^J ]\) if and only if the index itself is an absolute index.
In the case of the Shapley decomposition of the KBD index the result is the same. Assume that the EDE function is translatable so that $\Xi(y + \theta 1^N) = \Xi(y) + \theta$ and that each income source $j$ has $\theta^j 1^N$ added to it, then the contribution to measured inequality of each source is unchanged. To see this begin with the Shapley decomposition of $A(y)$,

$$A(y) = \sum_{j=1}^{J} [\mu^j - \sigma_j] \quad (2.62)$$

This formulation makes it easy to see that if $\sigma_j(y^1 + \theta^1 1^N, ..., y^j + \theta^j 1^N) = \sigma_j(y^1, ..., y^j) + \theta^j$, the Shapley decomposition is an absolute decomposition.

$$\sigma_j(y^1 + \theta^1 1^N, ..., y^j + \theta^j 1^N) =$$

$$\sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \sum_{|C|=c, C \subseteq \{j\}} \left[ \Xi(y^c + \theta^C) - \Xi(y^C + \theta^C - y^c - \theta^j) \right] \quad (2.63)$$

which equals

$$= \sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \sum_{|C|=c, C \subseteq \{j\}} \left[ \Xi(y^C) + \theta^C - \Xi(y^C - y^j) - \theta^C + \theta^j \right]. \quad (2.64)$$

Thus

$$= \sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \sum_{|C|=c, C \subseteq \{j\}} \left[ \Xi(y^C) - \Xi(y^C - y^j) + \theta^j \right], \quad (2.65)$$

which can be rewritten as

$$= \sigma_j(y^1, ..., y^j) + \sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \sum_{|C|=c, C \subseteq \{j\}} \theta_j. \quad (2.66)$$

Now recognising that there are $J-1$ choose $c-1$ coalitions of size $c$ which contain source $j$ as

$$= \sigma_j(y^1, ..., y^j) + \sum_{1 \leq c \leq J} \frac{(c-1)!(J-c)!}{J!} \frac{(J-1)!}{(c-1)!(J-c)!} \theta^j \quad (2.67)$$
This demonstrates that the Shapley decomposition of the absolute KBD index will be an absolute decomposition, no matter what the values of $\theta^j$ are.

### 2.5.2 Old vs New Decomposition Methods.

As I explained previously, a major objection to the use of the Fei, Ranis and Kuo decomposition is that the decomposition is not continuous at any point where there are ties in the ranking of overall income. For the decompositions that I have suggested, this is not a problem. The problem with the Fei, Ranis and Kuo decomposition is that the ranking of overall income determines the ranking of the income source. The decompositions presented in this paper are functions of the actual inequality index defined on the income sources and thus as long as the actual index is continuous in income, the decomposition is continuous in the component distributions.

An index that is homogenous of degree zero in incomes is said to be a relative index\(^\text{15}\). A desirable feature for any decomposition of a relative index is that each component of the decomposition is homogenous of degree 0 as well. Recalling the two decompositions of the AKS index

\[
I(y) = \sum_{j=1}^{J} s^j I(y^j) - \sum_{j=a}^{J} s^j [I(y) - I(y^j)]
\]

and

\[
I(y) = \sum_{j=1}^{J} s^j I(y^{-j}) + \sum_{j=1}^{J} s^j [I(y) - I(y^{-j})]
\]

If $I(y)$ is a relative index then so are $I(y^j)$ and $I(y^{-j})$. If each $y^j$ is multiplied by a constant $\lambda$ then the overall distribution is also multiplied by $\lambda$. Thus the numerator and the denominator of the shares $s^j$ are multiplied by $\lambda$ and the shares arc homogenous

\(^{15}\text{See Blackorby and Donaldson (1978) or Chakravarty (1990).}\)
of degree zero. Therefore the elements of the decompositions of the AKS index are 
homogenous of degree zero if and only if the overall index is homogenous of degree zero.

For the moment now consider only the first decomposition

\[ I(y) = \sum_{j=1}^{J} s^j I(y^j) + \sum_{j=a}^{J} [I(y) - I(y^j)] \]  

(2.71)

I will now show that this decomposition has, as two special cases, the decomposition of the 
Theil index that is presented in West and Theil (1991) and the standard decomposition of the 
coefficient of variation inequality index used by Layard and Zabalza (1979).

The Theil inequality index can be written as

\[ I^T(y) = \sum_{i=1}^{N} \frac{y_i}{N \mu} \log \left( \frac{y_i}{\mu} \right) \]  

(2.72)

and the decomposition of this index is given to be

\[ I^T(y) = \sum_{j=1}^{J} s^j \sum_{i=1}^{N} \frac{y_i^j}{N \mu^j} \log \left( \frac{y_i^j}{\mu^j} \right) - \sum_{j=1}^{J} \sum_{i=1}^{N} \frac{y_i^j}{N \mu} \log \left( \frac{y_i^j}{N y_i \mu^j} \right) \]  

(2.73)

Applying the first interactive decomposition technique for the AKS index gives an ex-
pression for the decomposition of

\[ I^T(y) = \sum_{j=1}^{J} s^j \sum_{i=1}^{N} \frac{y_i^j}{N \mu^j} \log \left( \frac{y_i^j}{\mu^j} \right) + C^I(y^1, ..., y^j, y) \]  

(2.74)

Comparing 2.73 to 2.74 and setting

\[ C^I(y^1, ..., y^j, y) = - \sum_{j=1}^{J} \sum_{i=1}^{N} \frac{y_i^j}{N \mu} \log \left( \frac{y_i^j \mu}{N y_i \mu^j} \right) \]  

(2.75)

shows that the two decompositions 2.73 and 2.74 are identical. Thus the decomposition 
presented in West and Theil (1991) is a special case of the decomposition presented in 
this paper.

Now consider the variance of income as an inequality index. The variance is

\[ I^V(y) = \sum_{i=1}^{N} \frac{(y_i - \mu)^2}{N} \]  

(2.76)
Chapter 2. Decomposing Inequality Indices by Income Source

<table>
<thead>
<tr>
<th>Row</th>
<th>Index</th>
<th>Decomposition</th>
<th>1</th>
<th>2</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gini</td>
<td>Interactive</td>
<td>1/12</td>
<td>1/12</td>
<td>-1/6</td>
</tr>
<tr>
<td>2</td>
<td>Gini</td>
<td>Shapley</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>r=1/2</td>
<td>Interactive I</td>
<td>0.018</td>
<td>0.0</td>
<td>-0.008</td>
</tr>
<tr>
<td>4</td>
<td>r=1/2</td>
<td>Shapley I</td>
<td>0.016</td>
<td>-0.014</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>r=1/2</td>
<td>Interactive A</td>
<td>0.05</td>
<td>0.0</td>
<td>-0.02</td>
</tr>
<tr>
<td>6</td>
<td>r=1/2</td>
<td>Shapley A</td>
<td>0.04</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>r=1/2</td>
<td>Interactive I</td>
<td>0.013</td>
<td>0.0</td>
<td>-0.008</td>
</tr>
<tr>
<td>8</td>
<td>r=1/2</td>
<td>Shapley I</td>
<td>0.01</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>r=1/2</td>
<td>Interactive A</td>
<td>0.05</td>
<td>0.0</td>
<td>-0.003</td>
</tr>
<tr>
<td>10</td>
<td>r=1/2</td>
<td>Shapley A</td>
<td>0.035</td>
<td>-0.01</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Amount of Measured Inequality Attributable to Different Sources

The standard decomposition of this inequality index is

$$I^V(y) = \sum_{j=1}^{J} \sum_{i=1}^{N} \frac{(y_i^j - \mu^j)^2}{N} + \sum_{j \neq k} \sum_{i=1}^{N} \frac{(y_i^j - \mu^j)(y_i^k - \mu^k)}{N}$$  \hspace{1cm} (2.77)

Applying the interactive decomposition technique for the KBD index yields

$$I^C(y) = \sum_{j=1}^{J} \frac{\mu^j}{\mu} \sum_{i=1}^{N} \frac{(y_i^j - \mu^j)^2}{N} + C^A(y^1, ..., y^J, y)$$  \hspace{1cm} (2.78)

Inspection of 2.77 and 2.78 reveals that the standard decomposition of the variance inequality index is a special case of one decomposition method explored in this paper.

It will be helpful in comparing my decompositions, both interactive and Shapley, with the Pseudo-Gini and Shorrock's decomposition, to calculate my decompositions for the examples used in section three as part of the justification for looking for different decomposition methods. Thus Table 2.1 gives the values of these decompositions for the three examples used in section three. The first two rows of table 2.1 give the decomposition of the Gini inequality index for $y^1 = [1, 2], y^2 = [2, 1]$, and $y = [3, 3]$.

As can be seen from rows one and two, the interactive decomposition of the Gini index in this case attributes the same direct effect on inequality to both income sources. Since the overall measured inequality is zero, the interaction effect shows that the interaction
between sources of income serves to reduce overall inequality. In contrast, the Shapley decomposition attributes zero to both income sources in this case. This demonstrates that both the interactive decomposition and the Shapley decomposition satisfy continuity in the vicinity of ties in the rank of overall income. This is because an individual's weight depends only on his or her rank in the income source $j$ for the interactive decomposition or his or her rank in coalition $C$ for the Shapley decomposition.

Rows three through six of table 2.5.2 give the values of the interactive and Shapley decompositions of the mean of order $1/2$ AKS index and KBD index. For the income distribution given by $y^1 = [1, \ 2], y^2 = [1, \ 1], \text{and } y = [2, \ 3]$. Since $y^2$ is equally distributed, both of the interactive decompositions show a direct effect of $y^2$ on measured inequality of zero. Both of these decompositions also show that there is a negative interaction effect. The Shapley decompositions show something different. The overall effect of source 2 is to reduce measured inequality while that of source 1 is to increase inequality. Thus source 2 has a non-zero contribution to inequality even though it itself is equally distributed. This obviously violates Shorrocks' axiom Normalization for Equal Factor Components.

Rows seven through ten of table 2.5.2 show the results of the mean of order $1/2$ decomposition of the distribution given by $y^1 = [1, \ 2], y^2 = [2, \ 2], \text{and } y = [3, \ 4]$. The interesting thing to notice in these results is the relationship between these results and the corresponding results in rows three to six. In the case of the interactive decomposition the contribution to inequality of source 1, and the interaction term have changed while the contribution of source 2 has not changed. Thus it would seem that the interactive decomposition suffers to some extent from the criticism that I levelled at Shorrocks' decomposition that in comparing the two distributions the only contribution not to change is the contribution of the income source that did change, at least for the relative AKS index. The Shapley decomposition does not suffer to the same extent since
the contribution of both sources changes.\footnote{This seems reasonable since the actual value of measured inequality changes from one case to another for all of the indices.}

One of the axioms that Shorrocks uses in his construction is the axiom of symmetry. This axiom states that the name of the income source does not affect the valuation of the contribution to income inequality. To see how this holds for the interactive decompositions of the AKS index, consider an economy with two sources of income. Consider now two income distributions $y$ and $\bar{y}$ which differ because $\bar{y}^1 = y^2$ and $\bar{y}^2 = y^1$. These two distributions have the same aggregate distribution so the only to have changed is the name of the income sources. Symmetry therefore requires $s^1I(y^1) + s^1[I(y) - I(y^1)] = \bar{s}^2I(\bar{y}^2) + \bar{s}^2[I(\bar{y}) - I(\bar{y}^2)]$ which is demonstrated by substituting in $y^1 = \bar{y}^2$. Similarly the second decomposition can be shown to be symmetric as well. The symmetry with respect to income sources is imposed on the Shapley decomposition as one of the axioms used in its characterization.

I have presented two plausible interactive decompositions of both the AKS and the KBD inequality indices, the first based on the distributions $y^j$ and the second based on the distributions $y^{-j}$. The two decomposition approaches provide slightly different information and thus are suitable in different applications. The first decomposition provides information about the degree of inequality within a given factor distribution and how the different factor component distributions interact to cancel out some inequality. The second decomposition shows the way in which the individual sources of income increase or decrease inequality in the aggregate distribution but does not say anything about the inequality within a factor component distribution. I also provided a method, based on the Shapley value, of exactly decomposing inequality indices which does not have an interaction term providing a much cleaner answer to how much inequality is a result of a given income source. In this section I have compared the different approaches to each
2.6 Potential Applications.

In this section I outline some possible applications of the decompositions of the inequality indices that I have presented. I also discuss the relation between the use of the decomposed inequality indices and the existing literature on these applications.

The first application is to applied taxation theory. Most people agree that progressivity is a desirable feature for a given tax system to have. A progressive tax system is one where the tax payments for an individual as a percentage of his or her income are increasing in that individual's income. A substantial literature has developed in measuring the degree of progressivity present in a given tax system\(^{17}\).

One observable feature of a progressive taxation scheme is that after-tax income is distributed more equally than before-tax income. This characteristic means that, given the same amount of revenue collected, the after-tax welfare\(^{18}\) of a society with a progressive taxation system is higher than the welfare of a society with a neutral or regressive taxation system\(^{19}\).

Define \(y\) to be after-tax income, \(y^1\) to be before-tax income and \(y^2\) to be tax payments. Then by definition \(y = y^1 - y^2\). Blackorby and Donaldson (1984) derive a relative index of global tax progression to be

\[
T^R(t, y^1) = \frac{I(y^1) - I(y)}{1 - I(y^1)} \tag{2.79}
\]

where \(t\) is the rate of neutral taxation that will produce the same level of revenue as the actual tax system. This index is zero if the taxation scheme is proportional, and is

\(^{17}\)See Blackorby and Donaldson (1984) or Pfingsten (1987) for example.

\(^{18}\)Given some degree of inequality aversion

\(^{19}\)A regressive tax system is one where relative payments decline with income and in a neutral system the relative payments are constant.
positive if the tax system is progressive.

Now consider the first decomposition of the AKS. In this case it is given by

$$I(y) = \sum_{j=1}^{J} s^j I(y^j) + \sum_{j=1}^{J} s^j [I(y) - I(y^j)]$$

(2.80)

The first interaction term contains the same information as the Blackorby and Donaldson index of tax progressivity. It is

$$s^j [I(y) - I(y^j)] = -s^j T^R(t, y^j) E(y^j)$$

(2.81)

Therefore the Blackorby–Donaldson index of tax progressivity has a strong connection with the kind of decomposition that I use in this paper. The decomposition, however, provides more quantitative information about the degree of equalization of income achieved by the tax system because it includes not just the progressivity of the tax system, but also information about the rest the components of income.

One possible problem with using this decomposition arises in the case where the tax system is a purely redistributive tax–transfer scheme. In this case the mean tax level is zero and the decomposition is undefined. In this situation it is necessary to separate the taxes and transfers and calculate the effects of the individual sources alone.

The second decomposition of the AKS index can also be used to produce the Blackorby–Donaldson relative index of tax progressivity. In this case it is the expression for the marginal contribution to inequality that is of interest.

Blackorby and Donaldson also provide an absolute index of tax progression that will provide some of the same information as the decompositions provided here.

Another area where the decomposition of income inequality indices is helpful is in development economics, in fact the Fei, Ranis and Kuo decomposition of the Gini was introduced as a tool in the economic development literature. Many theories of economic development such as Lewis (1954) predict that a particular pattern of inequality will
Chapter 2. Decomposing Inequality Indices by Income Source

occur over the development process. For example in the Lewis model income inequality is predicted to rise in the early stages of development as the income from capital increases and then fall as the supply curve for labour becomes more inelastic and the wage rate increases. Traditional tests of the Lewis model compare how inequality varies with mean income. An alternative to the traditional test of the Lewis model would require an income inequality index that can be decomposed by income source. If a time series of decompositions can be constructed then a rise in income inequality due to capital income followed by a decline in inequality from capital income in later periods would be support for the Lewis model. This test of the Lewis model makes use of a prediction that the traditional tests are unable to use.

The final application that I suggest is an investigation into the effect that human capital can have on income inequality. To the extent that earnings are increased by human capital, differences in education between individuals in a society will result in differences in income. This means that one possible method of reducing inequality is to change the distribution of education levels amongst the population. In determining the effectiveness of such policies, decomposable inequality indices are required to show exactly how the pattern of inequality changes when the distribution of education levels changes. A thorough examination of this question is left for the next chapter.

2.7 Conclusion

In this paper I have presented several ways of decomposing both the AKS inequality index and the KBD inequality index by income source. All of the decompositions share the same grounding in ethical inequality theory which allows for easy and intuitive interpretations of the individual terms of the decompositions. In contrast to most previous attempts to decompose by inequality source, I provide a general method that is appropriate for a.
wide class of inequality indices. I showed that several previously known decompositions of specific inequality indices can be considered special cases of the decompositions presented here. I concluded with a discussion of the possible uses of the decompositions in applied economics.
Chapter 3

Human Capital Models of Income Distribution
In this chapter I discuss the theoretical conditions which will ensure that the income decompositions of chapter four are correct. By correct I mean that not only does it make sense to speak of returns to education being part of earnings but also that the quantitative number that I determine for the returns to education is correct. For the earnings decomposition to be strictly valid two things have to be true; the first is that the human capital model of earnings must be an adequate explanation of earnings, the second is that the construction of the counterfactual incomes must capture all the general equilibrium effects of the change in the distribution of education. In this chapter I first discuss the main alternative to the human capital explanation of earnings, the market signalling hypothesis, and then illustrate, using a simple theoretical model, the arguments pertaining to whether a decomposition of the form I use is able to adequately reflect general equilibrium matters.

3.1 How Education Affects Earnings

The idea behind the human capital explanation of earnings is that people undertake education as an investment in future earning power. With more education an individual will have higher productivity at his or her job and thus will be paid a higher market wage. Any person who achieves a given level of education will have the same increase in productivity, and thus income, as any other person with the same personal characteristics. If the human capital model is an accurate description of the world then it makes sense that individual earnings can be separated into a part that is a return to education, and a part that is actually a return to other factors such as experience or gender.

The main competing hypothesis to the human capital hypothesis is the market signalling hypothesis of Spence (1973). This explanation maintains that education does not
necessarily contribute to increased productivity, although it may\(^1\), but that it does act as a signal to employers that the person will have high productivity at a given job. In the model of Spence (1973), individuals with high ability also have high marginal products and thus high incomes. The low ability individuals have lower productivity and a higher cost of education than the high ability individuals. Firms can use a high educational attainment as an indication that the individual is of high ability. Since the cost of achieving a given education level is lower for the high ability individuals, it is possible to set the pay for high ability people so that it is worthwhile for them to purchase education, but that it is not worthwhile for low ability types to purchase the education. The result is that only people with high ability will want to purchase education. Individuals will realize this and thus, when they are a high ability person, will purchase the education to indicate this fact to the employers.

The signalling model, like the human capital model, produces a high correlation between educational attainment and earnings but the regression of earnings on education does not yield the return to education. What this regression does yield is a measure of the return to ability. The reason that it cannot be considered a return to education is the possibility that, in the absence of education being used as a signal of ability, it is likely that another signal would be found. In this case no matter what the distribution of ability is, the same distribution of earnings would be observed. Constructing a counterfactual distribution of income as I do in this thesis will, in the case of the signalling hypothesis, result in two vectors of income. One is a return to other personal factors, as in the human capital model, but the other vector will not be a return to education, it will instead be the return to ability.

\(^1\)See Spence (1974), pp. 21, 22 for a model in which education does actually increase productivity but also acts as a signal of inherent ability.
As noted above, it is possible that education could provide both a signal of inherent ability and an actual increase in productivity. In this case the correlation between earnings and education is partly a result of signalling behaviour and partly a legitimate increase in productivity. This is not enough to save the construction of the counterfactual distributions unless the increase in productivity is much greater than the signalling effect in which case the signalling effect would be relatively unimportant. As the increase in productivity from education increases it becomes more likely to be so large relative to the cost of education that no signalling equilibrium exists. It is thus best to restrict attention to the pure human capital case.

For the analysis in the body of the thesis to be correct it is thus necessary that the human capital hypothesis be maintained throughout the analysis. Empirical tests of the human capital model which contain separate measures of ability, such as IQ scores, have been implemented in the literature. These studies generally show that education, even after controlling for the effect of ability, is a significant determinant of earnings. In addition Albrecht (1981) in a direct test of the signalling model is unable to find support for the signalling explanation of the correlation between education and productivity. These results at least partially support the human capital hypothesis. It is also true that many high paying jobs do require a certain level of education, regardless of the individuals ability. This type of evidence also tends to support the human capital model of earnings. I therefore adopt the human capital model as a reasonable theoretical model on which to pin the rest of the analysis in the thesis.

---

2See for example Taubman (1975), Taubman and Wales (1974)
3Of course not supporting the signalling model is not the same as supporting the human capital model but Albrecht's results do also show a positive correlation between education and productivity.
4For example medical doctor or engineer.
3.2 General Equilibrium Issues

The second major condition that must hold for the decomposition analysis to be valid is that the construction of the counterfactual income levels must contain all of the general equilibrium effects on earnings of a change in the distribution of education. In general, we can expect that changing the distribution of the educational attainment throughout society will change the distribution of returns to education. As the supply of specific skills change so will the returns to these skills that are generated in the market. Thus changing the distribution of skills will change not only the level of returns to education, it will also change the relative returns to different types of education. I explain in this section the theoretical conditions under which these effects on the distribution are unimportant in the analysis. I then outline a reasonable model for which evidence on the direction of bias exists. This section draws heavily from Lucas (1977).

Suppose that earnings for person $i$ vary according to the function

$$y_i = \beta_i + r_i a_i + e_i$$  \hspace{1cm} (3.82)

Where $y_i$ is person $i$'s income, $r_i$ is the return to person $i$'s education, $a_i$ is the amount of person $i$'s education, $e_i$ is a random error term, and $\beta_i$ is an individual specific intercept, intended to capture the effect of personal characteristics, other than education on the individual's earnings. A human capital earnings equation, run on cross sectional data will result in the following predictive equation

$$y_i = \hat{\beta} + \hat{r} a_i + u_i$$ \hspace{1cm} (3.83)

Equation 3.83 has the following interpretation. $\hat{\beta}$ is the average level of returns to other factors than education, $\hat{r}$ is the average return to education within the sample and $u_i$ is the residual.
Chapter 3. Human Capital Models of Income Distribution

Equation 3.83 is a linear equation so the expression for the variance of income as a function of the variance of the other terms is easy to determine. Suppose equation 3.83 is used to calculate the variance of income\(^5\) then the resulting expression for the variance of predicted income is

\[
\sigma(y) = \hat{r}^2 \sigma(a)
\]  
(3.84)

where in general \(\sigma(y)\) is the variance of the predicted income and \(\sigma(a)\) is the variance of the education variable. Corresponding to this is the expression for the variance of actual income, obtained from the linear equation 3.82,

\[
\sigma(y) = \sigma(\beta)\hat{r}^2 \sigma(a) + \hat{a}^2 \sigma(r) + (\sigma(a)\sigma(r))^{\frac{1}{2}} + \sigma(\epsilon)
\]  
(3.85)

where \(\sigma(y)\) is the variance of \(y_i\), \(\sigma(\beta)\) is the variance of \(\beta_i\), \(\sigma(r)\) is the variance of returns to education \(r_i\), and \(\sigma(a)\) is the variance of educational attainment. Thus I now have an expression for the variance of predicted income and another expression for the variance of actual income, relating both to the variance of actual educational attainment.

The next step is to see how these two expressions vary with the distribution of the education variable. Accordingly assume that the variance of \(a\) changes but the mean of the distribution of \(a\) does not change\(^6\). The change in the variance of predicted income is therefore calculated as the derivative of equation 3.84 with respect to \(\sigma(a)\), which is

\[
\frac{\partial \sigma(y)}{\partial \sigma(a)} = \hat{r}^2
\]  
(3.86)

and the change in the variance of actual income is

\[
\frac{\partial \sigma(y)}{\partial \sigma(a)} = \hat{r}^2 + 2\sigma(a)\hat{r}\left(\frac{\partial \hat{r}}{\partial \sigma(a)}\right) + [\hat{a}^2 + \sigma(a)]\left(\frac{\partial \sigma(r)}{\partial \sigma(a)}\right) + \sigma(r)
\]  
(3.87)

---

\(^5\)The variance of income is used in this subsection because it makes the exposition much easier. The lessons learned using the variance apply equally well to other measures of inequality such as the AKS S-Gini index.

\(^6\)This differs from the construction of the counterfactual distribution in the body of the thesis because there I change both the mean education level and the variance of education. This would not affect the calculation of the variance since the variance is not mean–dependant.
Constructing the counterfactual distributions the way I do and then analyzing how the inequality changes when the counterfactual distributions are used is analogous to using equation 3.86 instead of equation 3.87 to determine how a change in the distribution of education will affect the distribution of earnings. With equations 3.86, and 3.87 the bias inherent in using equation 3.86 to estimate the effect of a change in education on the change in the distribution of income is

$$\mathbf{B} = \frac{\partial \sigma(y)}{\partial \sigma(a)} - \frac{\partial \sigma(y)}{\partial \sigma(a)} - \left[ 2\sigma(a)\bar{F} \left( \frac{\partial \bar{r}}{\partial \sigma(a)} \right) + \left[ \bar{a}^2 + \sigma(a) \right] \left( \frac{\partial \sigma(r)}{\partial \sigma(a)} \right) + \sigma(r) \right] $$  (3.88)

From equation 3.88 it is apparent that the degree of bias in the estimation of the effect of the change in the distribution of education is dependant on the sign and magnitude of the expressions

$$\frac{\partial \bar{r}}{\partial \sigma(a)}$$  (3.89)

and

$$\frac{\partial \sigma(r)}{\partial \sigma(a)}$$  (3.90)

A special case deserves some mention. It is when the return to education is the same for all individuals. When this is the case, $\sigma(r)$ is zero, and both expression 3.89 and 3.90 are zero. In this case the bias is zero.

Only a very restricted form of an economy production function will result in a bias of zero in equation 3.88. For the bias to be zero the schooling of the workers must enter into the economy production function separately from the supply of labour. Lucas (1977) demonstrates that an economy production function of the following sort is the only type which will generate a bias of zero. Suppose the economy production function is

$$Q = F(K, L, A)$$  (3.91)

where $Q$ is the quantity of an aggregate output, $K$ is the aggregate capital input, $L$ is the quantity of raw labour, and $A$ is the sum of schooling over all individuals in the economy.
In this case the amount of schooling enters separately into the production function. An economy such as this will have everyone with the same return to education. In this case the variance of the return to education is zero and everyone in the economy has the mean education level. In this case the bias is zero. In any other case the bias introduced by using the estimated human capital earnings equation is most likely non-zero.

Since the technology that generates a zero bias term is so restrictive, it is unlikely to hold in practice and I must accept the likelihood that there is a non-zero bias in the construction of the counterfactual distributions.

Given a non-zero bias it is worthwhile to try and analyse the terms of the bias to attempt to see what signs for the bias may be reasonable. As an illustration of a reasonable case where the sign of the bias can be determined consider the case of an aggregate production function for an economy of two people\(^7\) such as

\[
Q = F(a_1l_1, a_2l_2)
\]  

(3.92)

where, without loss of generality, \(a_1 > a_2\). \(a_1\) is the amount of education of person 1, \(l_1\) is the labour supply of person 1, with similar notation for person 2. Assume that the aggregate production function is symmetric with respect to labour efficiency units and is strictly increasing and concave. Given these assumptions

\[
r_1 = \frac{\partial Q}{\partial a_1} = l_1F_1(a_1l_1, a_2l_2) > 0
\]

(3.93)

and

\[
r_2 = \frac{\partial Q}{\partial a_2} = l_2F_2(a_1l_1, a_2l_2) > 0
\]

(3.94)

where \(r_1\) and \(r_2\) are the returns to education for individuals 1 and 2 respectively. These two expressions can be combined to yield the average return to education in the economy,

\[
\bar{r} = \frac{l_1F_1(a_1l_1, a_2l_2) + l_2F_2(a_1l_1, a_2l_2)}{2}
\]

(3.95)

\(^7\)Alternatively two types of people, skilled and unskilled.
Now an increase in the variability of education \( a \), given these assumptions is equivalent to an increase in \( a_1 \) and a decrease in \( a_2 \). To keep the mean education level the same, the amount added to \( a_1 \) must be the same as the amount subtracted from \( a_2 \). The total differential of \( \bar{r} \) for the changes to \( a_1 \) and \( a_2 \) is therefore

\[
d\bar{r} = \frac{1}{2}(l_1^2 F_{11}(a_1l_1, a_2l_2) + l_1l_2 F_{21}(a_1l_1, a_2l_2))da_1
\]

\[
+ [l_2^2 F_{22}(a_1l_1, a_2l_2) + l_1l_2 F_{12}(a_1l_1, a_2l_2)]da_2
\]

which since \( da_1 = -da_2 \), is equal to

\[
d\bar{r} = [l_1^2 F_{11}(a_1l_1, a_2l_2) - l_2^2 F_{22}(a_1l_1, a_2l_2)]de/2
\]

where \( de \) is the common absolute value of \( da_1 \) and \( da_2 \). Since the production function is symmetric in efficiency units, the same number of efficiency units will be hired from each type, that is \( a_1l_1 = a_2l_2 \). This implies that \( F_{11}(a_1l_1, a_2l_2) = F_{22}(a_1l_1, a_2l_2) \) and

\[
\text{sign} d\bar{r} = -\text{sign}(l_1^2 - l_2^2) > 0
\]

Thus a good case can be made that the average return to education is increasing in the variance of the distribution of education. This occurs because concavity of the production function implies that the decrease in the return to education for person 1 is of a smaller absolute value than the increase in the return to education of person 2. It remains to be shown how the second moment of the distribution of returns to education varies with the variance of the distribution of education.

Recall that the return to education for the first type of individual is given by the following expression

\[
r_1 = l_1 F_1(a_1l_1, a_2l_2)
\]

and the derivative of \( r_1 \) with respect to \( a_1 \) is

\[
\frac{\partial r_1}{\partial a_1} = l_1^2 F_{11}(a_1l_1, a_2l_2) < 0
\]
Since \( r_1 \) is a monotonic function of \( a_1 \), an increase in \( a_1 \) will cause a corresponding decrease in the value of \( r_1 \). Similarly a decrease in the value of \( a_2 \) will result in an increase in the value of \( r_2 \). Thus an increase in the value of the variance of \( a \) will result in a corresponding increase in the variance of \( r \), so that

\[
\frac{\partial \sigma(r)}{\partial \sigma(a)} > 0
\]  
(3.101)

The results so far indicate that, in the special case above, \( B \), the bias inherent in ignoring the general equilibrium effects is negative. This means that the inequality of the resulting predicted distribution of income is greater than the corresponding result for the actual income. The way that this affects the decomposition of earnings in the body of the paper is that the variability of the base returns in my constructed decomposition of income will be greater than the true variance of base returns. Since the distribution of actual earnings does not change, an increase in the variability of the base returns results in a decrease in the variability of the estimated returns to education, relative to the true distribution of returns to education.

The above has made a case for being able to sign the bias resulting from ignoring the general equilibrium effects when constructing the counterfactual distributions. Even the example that I presented, where production depends on efficiency units has some heroic assumptions. The first and most important being whether the efficiency units formulation of the aggregate production function is an acceptable one. To begin with there is the question of whether or not an aggregate production function exists at all. If the aggregate production function does exist then the next problem is whether the efficiency units formulation is appropriate. As some authors have pointed out\(^8\) the efficiency units hypothesis implies that raw labour and education are perfect substitutes. We can achieve the same output with a lot of labour and little education or with a lot of education and

\(^8\)See Lucas (1977)
little labour. For many occupations, such as labourer, this is probably an inadequate
description of reality.

An important assumption in the theoretical model of this chapter that is certainly
not true in my empirical model is that the return to education is independant of other
personal characteristics. In the empirical model in the other chapters I explicitly assume
that the return to education interacts with experience and gender. This means that
even if the other assumptions are correct in the analysis signing the bias, there is still
some uncertainty about it’s applicability to signing the bias in my analysis of returns to
education

Even given the probability that the analysis above, that signs the bias, is not directly
applicable to my situation, the exercise in this thesis is still worth doing. There is very
little econometric work done that is truly general equilibrium, everything isolates some
subset of the variables that do affect a particular dependent variable because a true
general equilibrium model is impossible to estimate and a partial equilibrium model is
better than nothing at all. Thus even if the signing of the bias above is not accurate I
justify the analysis presented in this thesis by saying that the question is an important
one to analyze from a social perspective and when a true general equilibrium model is
unavailable, then a partial equilibrium model is the next best alternative.
Chapter 4

The Contribution of Education to Earnings Inequality
4.1 Introduction.

Empirical evidence and economic theory both indicate that education is an important influence on an individual's expected earnings. An implication of this is that differences in educational attainment have an important influence on differences in labour earnings. This means that social policies which affect the education system, such as funding cutbacks, can have important effects on the distribution of income in later time periods. Much of the current public discussion about equality of access to education is predicated on the assumption that restricting access to education for individuals with low earnings will result in a perpetuation of earnings inequality from generation to generation. Models such as Loury (1981), Becker (1964), and Spence (1973) provide a theoretical justification for these beliefs. Empirical work has examined in great detail the connection between education and average earnings but very little analysis has been done of the effect of the distribution of education on the distribution of earnings. Since society cares, not only about the level of national income, but also about the distribution of income, an important policy question has been largely ignored. In this chapter I present an analysis of the contribution of differences in education levels to earnings inequality in Canada.

The second chapter of this thesis examined the decomposition of income inequality measures by income source; thus the next step in the empirical analysis is to decompose earnings itself into factor components. I draw on the empirical labour economics literature to estimate earnings equations using a sample drawn from the 1986 Survey of Family Expenditures compiled by Statistics Canada. These estimated equations are then used to determine how the earnings distribution would change if everyone had the same base level of education. The resulting distribution is used to decompose earnings into the base level and an amount attributable to education. First I consider what would happen if an individual's actual education level was replaced with a counterfactual education level.
The estimated earnings equations are used to predict what the resulting earnings would be if the individual had this counterfactual level of education. The difference between his or her current earnings and the counterfactual earnings level is the contribution of education to the person's earnings.

The estimated vectors of personal earnings are used in a decomposition of the S-Gini index of relative inequality to determine the contribution of education to measured earnings inequality. The S-Gini index allows different degrees of inequality aversion to be used in the measurement of inequality and this flexibility is one of the major differences between this work and the previous literature. The earnings decomposition is first performed ignoring any effects that education may have on the probability of being employed and then again considering the change in the probability of working as part of the return to education. The inclusion of this second effect is another of the distinguishing features of this analysis compared to previous work on the same subject. I further examine the differing effects by level of education to see if there is a difference between the effect of secondary education and post-secondary education. The last part of the chapter is a comparison of the results using the different methods, introduced in the second chapter, of decomposing the inequality index.

The question of the contribution of differences in education to earnings inequality has been addressed empirically in previous work, notably Taubman (1975) and Layard and Zabalza (1979). A major difference between my work and theirs is that I use the explicitly ethical decomposition of the S-Gini inequality index introduced in the previous paper. This ethical approach allows a numerically meaningful measure of the effect of education on earnings inequality and yields both different results, and a different interpretation of these results than the analysis done by Taubman as well as Layard and Zabalza.

The rest of the chapter is as follows. In section two I discuss three important issues that must be dealt with in any attempt to determine the contribution of education to
earnings inequality. In section three I discuss the previous empirical studies of Taubman (1975) and Layard and Zabalza (1979). The techniques used by the previous studies are applied to my data in section four to provide a reference point for the later ethical analysis. In section five I estimate the earnings equations and construct the counterfactual earnings distributions used in the inequality analysis. Section six presents the estimated contribution of the two sources of earnings to measured earnings inequality for two methods of determining the contribution of education to earnings. In section seven I examine how the contribution of education to earnings inequality varies with education level. Section eight presents a comparison of the results from the direct interactive decomposition with the Shapley decomposition. Section nine concludes.

4.2 Research Strategy

In this section of the paper I outline the three major problems which should be dealt with in an examination of the contribution of returns to education to earnings inequality. The first problem is a question of what is the appropriate index to use in measuring inequality, the second is how to measure the effect of one source of income on the overall inequality, the last is how to properly decompose earnings into a part due to education and a part due to other factors.

The first consideration that must be made in an attempt to quantify the effect of education on earnings inequality is how to measure earnings inequality. There are basically two general methods of doing this. The first are the statistical methods which involve estimating the variance of earnings or some other measure of the dispersion of the earnings distribution. This type of analysis is useful in a purely descriptive exercise where the objective is to describe an income distribution, or compare two distributions, but not to determine their desirability. The other general approach to the measurement
Chapter 4. *The Contribution of Education to Earnings Inequality*

of inequality is the *ethical* approach. In the ethical approach the first step is to decide on a social evaluation function which incorporates the ethical values of society in regards to income inequality. Then one can construct a suitable inequality index from the data using a transformation of the social evaluation function.

It is my contention that the appropriate way to measure inequality in this paper is by the ethical approach. I make this claim for two closely related reasons. The first being that, as a society, it is social welfare that we are ultimately concerned with and earnings inequality is part of the social welfare and our social ethics must be included in our measure of inequality. Thus if one wants to make any policy recommendations, one must use an ethical index. The statistical indices are just descriptions of the earnings distribution with no way to determine whether one value is better or worse than another. Any attempt to rank statistical indices must rely on ethical considerations and then the ethical judgements should be made explicitly instead of hiding them behind a veneer of objectivity. The ethical index I use and its implications for the social preferences are outlined in section six.

How to determine the effect of education on measured inequality is simply a question of how best to decompose an earnings inequality index by earnings source. To do this properly for an ethical index of inequality requires that the decomposition itself be ethical and thus includes the explicit ethical judgements of the society as well as having the interpretation of the loss of welfare in the current distribution as a result of income inequality. Without using the ethical approach to decomposition, there is no natural way of selecting a decomposition rule from the infinity of possible decompositions. Insisting on an ethical approach does provide a natural way to select an appropriate decomposition rule. Appropriate methods for doing this were suggested in the previous chapter of the thesis and are briefly outlined in section six of this chapter.

The question of how best to measure the returns to education in the context of
determining earnings inequality arises because at any given time the population with which I am concerned consists of two groups, the employed and the unemployed. The unemployed will have earnings of zero while the employed will have some positive value of earnings. This has a very strong implication for the shape of the earnings distribution. Even with a continuous distribution of potential earnings, an individual will have a strictly positive probability of a realization of zero earnings, due to the possibility of being unemployed. The density function of observed earnings then has a mass point of positive probability at zero earnings and some continuous distribution over the positive part of the real line. In examining how the distribution of education affects the distribution of earnings the effect of education on the mass point at zero must explicitly be considered. If, for example, increased education results in an increased probability of being employed, then part of the return to education must be this increase in the probability that an individual will be employed. If this effect is not included in the return to education, then certainly the share of earnings that are due to education will be underestimated and to the extent also that the amount of underestimation of the return to education is correlated with education level, this will affect the distribution of the return to education. Both of these possibilities have substantial effects on the amount of earnings inequality that can be attributed to education.

A typical distribution that might arise in this context is demonstrated in figure 1. This distribution may arise from the following process. Desired labour supply is determined by the equation

\[ h_i^* = g(a_i, x_i) + \nu_i, \]

(4.102)

where \( a_i \) is the education level of person \( i \), \( x_i \) is a vector of non-education personal characteristics, and \( \nu_i \) is a mean zero, homoskedastic error term. Positive hours are worked if and only if \( h_i^* > 0 \) and then \( h_i = h_i^* \), actual hours worked equal desired hours.
Chapter 4. The Contribution of Education to Earnings Inequality

If $h_i^* \leq 0$ then zero hours are worked, $h_i = 0$. If and only if hours are positive, an individual receives earnings determined by the equation

$$y_i = y_i^* = f(a_i, x_i) + \epsilon_i,$$

(4.103)

where again $\epsilon_i$ is a mean zero, homoskedastic error. If hours are zero then $y_i = 0$. The measurement of the contribution of a given set of personal characteristics to earnings involves two steps. The first step is to estimate the parameters of the earnings generating process, the second is to use these estimated parameters to decompose earnings. The censoring in the observed earnings distribution creates a problem in both the estimation of the earnings process, because of the need to estimate both the mass point and the continuous part of the underlying density, and in the decomposition of observed earnings into a part due education and another part due to other factors.

The first step, the estimation of the earnings process is a step well understood by labour economists. A problem arises because of the likely correlation between the error terms $\nu$ and $\epsilon$. This correlation, if it is non-zero results in a non-zero expected error term in the regression determining the parameters of $f$. The result of the non-zero expected error term is a misspecification bias of the estimated parameters for the function $f$.

Three methods can be considered for handling the presence of the mass point and the subsequent estimation of the function $f$. The first possible method is to completely ignore the observations with zero earnings and to estimate $f$ with only the non-zero observations. The omission of the zero earnings obviously determines nothing at all about the mass point, in addition the shape of the estimated earnings distribution is not correct. If figure 1 shows the true shape of the observed earnings distribution, and if $\epsilon$ and $\nu$ are positively correlated (a reasonable situation), then the estimated distribution of $y$ will be like figure 2. In this the estimated distribution is skewed to the right, compared to the actual distribution. This will have the result that the mean of the distribution is
larger, the left tail of the continuous portion of the distribution is smaller and the right tail is larger. This is in addition to the errors introduced because the mass point is not used at all. Both the errors in the estimation of the continuous portion and the omission of the mass point will lead to underestimation of earnings inequality with an index with any amount of inequality aversion at all.

The next possibility is that the zero earnings observations are used but with no recognition that they are part of a mass point or that the process generating a zero is in any way different from the process generating a positive $y$. In this case the resultant estimated distribution will resemble figure 3. The mean of the distribution will be close to the mean of the censored distribution but the estimated shape is very different. The mass point will not be present and the distribution is substantially flatter than the censored distribution. It is not clear exactly how this will affect measured inequality but it will affect the results.

The third, and in this situation correct, way of estimating the earnings process is to jointly estimate the probability of a zero earnings observation, conditional on $a$ and $x$, and the parameters of the function $f$.

I have so far established that the estimation of earnings must take into account both the spike of individuals earning zero and the rest of the density of possible earnings. There are several ways of incorporating the possibility of some individuals working zero hours into a model of earnings$^1$. I will concentrate on four; the simple tobit specification, the fixed cost of working specification, the minimum hours of work specification, and a two stage tobit model. All four approaches concentrate on the relationship between the decision to participate in the labour market and the decision on the number of hours that are worked.

The simple tobit specification postulates that desired hours of work are determined

$^1$See for example Zabel (1993) or Heckman and Macurdy (1986)
by an equation which depends on personal characteristics. If desired hours of work are positive then the individual enters the workforce and works the desired hours. If the desired hours are zero or negative the individual does not enter the workforce. Since the same reduced form equation determines both the hours of work decision and the participation decision, the effect of a given characteristic on the participation decision must be a proportion (constant across characteristics) of the effect on hours of work. There is evidence\(^2\) that this model tends to overstate the effect of wages on labour hours. It remains a popular choice of model for applied work because it is relatively simple to implement.

The second model I consider is the fixed cost model of Cogan (1981). In this model the worker is assumed to undergo some fixed costs of entering the labour force. The effect of this is that the worker will not enter into the labour force unless a minimum number of hours will be worked, any less and the benefit from working does not cover the fixed cost of entering the labour force. Empirically this model loosens the restriction on the effect of an individual's characteristics on the labour supply and participation decision that is present in the simple tobit specification.

The third model to be considered is the minimum hours model of Moffit (1982). This model recognises that the number of hours worked is a result of two factors, demand for labour, and supply of labour. The demand for labour is often institutionalized, especially in regards to hours of work. For example it is very difficult to find a job where the employer will allow an employee to work only half an hour a week. The minimum hours model has employers insisting that employees work a minimum number of hours if at all. A potential worker whose desired hours are above the minimum is not constrained in his or her choice but a worker who's desired hours fall between zero and the minimum faces the choice to either work the minimum or not work at all. This model provides

\(^2\)See Mroz (1987).
another way of loosening the restriction imposed on the effect of personal characteristics on the labour supply decision and the participation. It does however greatly increase the complexity of the model.

The last model to be considered is a two stage tobit specification. In the first stage a decision is made about whether to participate in the labour force or not. In the second stage the desired hours are determined. The actual number of hours worked depends on the results of both of the equations. Since different equations determine participation and labour supply this model also breaks the connection between the labour supply and participation decisions that is imposed in the simple tobit. Again the drawback is the increased complexity of the estimation.

The choice between the models then is one where increasing flexibility is traded off with increasing difficulty. In my case this increased difficulty is doubly important since the major purpose of estimating the earnings is the construction of counterfactual earnings distributions when education levels are changed. In the case of the simple tobit specification this is relatively straightforward, anyone with predicted negative earnings does not work, while anyone with positive predicted earnings does work\(^3\). With the next two models the potential for fixed effects or minimum hours would greatly complicate the construction of counterfactual earnings. The two–stage tobit specification adds a further complexity in the interaction between the hours of work equation and the labour force participation.

Thus simplicity is the overriding factor in the choice of earnings model. I get this simplicity with some cost however. As discussed above the simple tobit model tends to overstate the effect of wages on hours of labour supplied. This may have an impact on my counterfactual distributions. Decreased education tends to produce lower wages. The overstatement of the elasticity of labour supply could mean that, for those not driven out

\(^3\)This is covered more completely in later sections.
of the labour force, the decrease in earnings may be overstated. On the other hand the participation decision should not be adversely affected\(^4\). The simple tobit specification of earnings is dealt with more formally in section five.

Although the above discussion has been about labour supply models, in what follows below I do not actually estimate labour supply, I instead estimate an earnings equation. When I estimate an earnings equation the dependant variable is the product of labour supply and wages. Thus a change in the observed earnings could be a result of a change in hours, a change in the wage rate, or a combination of the two. The parameters of the estimated equation thus serve two purposes, the first is to capture the effect of personal characteristics on labour supply, the second is to capture the effect of the same characteristic on wages. The fact that one parameter is asked to do two things means that the model is a compromise. Again this compromise is imposed for tractability, but my opinion is that it is likely to be relatively unimportant to the analysis.

It is appropriate to justify the use of a particular specification of earnings by appealing to labour supply models because most of the differences between the different labour supply models arise in the treatment of the participation decision, or the transition from work to non-work. With an earnings equation this decision shows up in the observation of zero or positive earnings. Because earnings are the product of hours and wages zero earnings can arise either from zero hours, in which case the above models of labour supply are appropriate, or zero wage. Zero wage is very unlikely\(^5\) so treating observations of zero earnings as non-participation decisions and using labour supply models to analyze them seems justified.

The second step in attributing a part of earnings to education is to use the estimated

---

\(^4\)Zabel (1993) does find that the simple tobit performs marginally worse at this task than the other models. The difference however is about half of a percentage point in the probability of correct predictions which is around 80% for all models.

\(^5\)especially with minimum wage laws.
earnings process to separate the effect of education from the other variables. Because the probability of each individual being in the mass point at zero earnings is determined in part by the desired hours function $g$ which is a function of education, the mass point must be allowed to change when the counterfactual education levels are constructed. Thus the size of the mass point in the observed distribution is determined in part by the distribution of education.

So far in this section I have been referring to the non–workers as a homogenous group. In fact this is not the case. The non–workers actually consist of two distinct types of individuals. The first group is those individuals who are not in the labour force, ie not working and not looking for work, while the second group are those people who are in the labour force but cannot find work, ie the unemployed. The data that I use make no distinction between the unemployed and those not in the labour force and so I am forced to treat these two groups as one. Since I consider both heads of households and spouses in this thesis it is likely that both the unemployed and not in labour force are well represented in the sample. If the two groups react differently to labour market stimuli then the results of the study may be affected. Recent work by Gonul (1992) suggests that for men out of the labour force and unemployed are not distinct states but for women, individuals in the two groups behave differently. Thus the homogenous treatment of the two groups is likely to be a limitation.

### 4.3 Previous Empirical Analyses.

In this section I present a brief review of two previous studies of the effect of education on earnings inequality. There have been a number of theoretical studies of how education affects earnings, and thus earnings inequality, beginning with Becker (1964). The actual effect of education on earnings inequality is an empirical issue, consequently I confine
myself to the applied work that has been done on this question.

A major study of the effect of education on earnings inequality was undertaken by Taubman (1975), building on work done by Taubman and Wales (1974). In this book Taubman is concerned with determining the existence and shape of the wage earnings distribution when various sources of earnings are removed from the overall distribution. The primary objective is to determine from which probability distribution the wage earnings is being drawn. Thus Taubman does not consider a normative welfare analysis of the earnings distribution.

The first step in Taubman’s analysis is the estimation of a human capital earnings equation from a sample of individuals who entered the U.S. Army Air Cadet program during 1943. The data on educational attainment, personal characteristics and job experience was collected by survey in 1955 and 1970. This provides a rich data source on individual characteristics, including a measure of ability, as well as data on job experiences but, because of the fairly stringent conditions that had to be met to enter into the Air Cadet program, the data is not a random sample of the U.S. population.

The earnings equation estimated by Taubman is of the form

\[ y = X\beta + u \]  \hspace{1cm} (4.104)

where \( y \) is labour earnings, \( X \) is a vector of personal characteristics, \( \beta \) is a vector of parameters to be estimated and \( u \) is an error term. Included in the vector \( X \) is education level, a measure of ability, some indications of family background and occupation. This equation is estimated twice, once for each year in which the survey was undertaken.

With the estimated earnings equations, Taubman proceeds to separate the effect of various subsets of the explanatory variables on the earnings distribution. The main experiment with which I am concerned is the one where the effect of education is removed. Define \( y_2 = y - X_2\beta_2 \), where \( X_2 \) is the set of education variables, to be the earnings with
the effect of education removed. Taubman then compares the inequality in $y_2$ with the inequality in total earnings $y$.

Taubman uses the standard deviation as his measure of inequality. This has the bad feature that it is not invariant to scale; a change in the units of account will change the amount of measured inequality. The coefficient of variation, which is the ratio of the standard deviation of earnings to the mean of earnings is scale invariant and is another of the measures of inequality used by Taubman. In 1955 the mean earnings in the sample was $7300 and the standard deviation was $3800 for a coefficient of variation of 0.52. In 1969 the mean earnings was $14500 with standard deviation $9400 for a coefficient of variation of 0.65. All earnings are calculated using 1958 prices.

For 1955 the standard deviation of the non education–related earnings, $y_2$, was $3740 compared to $3810 for total earnings. In 1969 the standard deviation of non–education earnings was 9110 compared to 9420 for total earnings. Thus if all effects of education were removed, the standard deviation of earnings would have fallen in 1955 by $70 and in 1969 it would have fallen by $310. This indicates that differences in educational attainment increase the variance of the earnings distribution. The interpretation of these results must be influenced by the recollection that the sample is restricted and the extension of these results to the general population must be done with caution.

A possible objection to these results is that the specification of the earnings equation is very restrictive and allows no interaction to occur between the different independent variables. A more general earnings equation may produce differing results. Another problem arises because only the variance of earnings is reported for $y_2$. The variance of earnings is an absolute measure of inequality and as such is not changed by a change in the earnings distribution of the form $\hat{y} = y + \alpha 1^N$, for all $\alpha$. It will change if the incomes are all multiplied by a constant, such as happens when Taubman converts all earnings in 1955 and 1969 to real earnings in 1958 dollars. His results are sensitive to the year that
he chooses\textsuperscript{\textit{6}}.

The second major study of the effect of education on the distribution of earnings that I consider is Layard and Zabalza (1979). In this paper Layard and Zabalza attempt to use family earnings in their analysis of earnings inequality. They use two related measures of inequality, the coefficient of variation and the squared coefficient of variation. The data used in this study is from the general social survey for 1975 and consists of 4027 British households.

The overall inequality index for family earnings is 0.26 where the index is the squared coefficient of variation. To decompose family earnings by factor source, Layard and Zabalza first estimate the equation

$$f = \alpha_1 + \alpha_2 s_1 + \alpha_3 s_2 + \alpha_4 p_1 + \alpha_5 p_2 + \alpha_6 X_1^* + \alpha_7 X_2^* + v$$

(4.105)

where $s_1$ is the age at which formal schooling ended for the male, $p_1$ is the profession of the male and $X_1^*$ is an experience variable for the male. Variables with a subscripted 2 pertain to the female. Using the estimated equation, the contribution of the education variables to the squared coefficient of variation of family earnings is 0.052 compared to an overall value of 0.26. Thus about 20\% of total family earnings inequality can be explained by differences in education.

Layard and Zabalza continue with some analysis of possible policy changes using their model. The first policy they consider is an increase in the quality of education represented by an increase in per pupil expenditure. Assuming rates of return on schooling expenditure of 3\% and 6\%, this policy results in a less unequal earnings distribution. Raising the age of compulsory education serves to increase earnings inequality slightly. Grants to those individuals who stay in school serve to increase inequality in earnings.

\textsuperscript{6}It is possible, given the price index used, to determine how Taubman's results are affected by this scaling but a different price index will change the results of his analysis.
Neither study accounts for any interaction between education and other personal characteristics, such as job experience. This is likely to have significant impacts on the results, as it forces the experience profile of earnings to be identical across education levels.

4.4 Initial Comparisons

In this section I replicate the techniques used in Taubman (1975) and Layard and Zabalza (1979). The results can be used to give context to the later results and to provide some information about how much of the difference between my results and these previous studies are attributable solely to the different data.

The data that I use are drawn from the 1986 Survey of Family Expenditures compiled by Statistics Canada. The personal characteristics of the head of the household and the spouse were selected and individuals not reporting an education level, and those over the age of sixty-five were cut from the sample. It may be more traditional to consider only full time workers when doing the estimation. I do not do this since I am concerned, not only with the workers in society, but also with the non-workers. Using a sample of only workers would result in selection bias being a problem. Further details are provided in the data appendix. Detailed results for each of the regressions run in this section are provided in appendix two.

In the study done by Taubman, the first step is to estimate an earnings equation of the form

\[ y = X\beta + \epsilon \]  \hfill (4.106)

where \( X \) is a vector of personal characteristics, \( y \) is the level of earnings, and \( \epsilon \) is a random error term. I estimate a version of this equation for my own sample where the vector of personal characteristics contains education level, the experience measure, marital status,
geographic location, sex and first language. I do not include any interactions in this equation. This is as close to Taubman’s equation as I am able to get with my data. The equation I estimate is therefore

\[ y = \alpha_0 + \alpha_1 ed_i + \alpha_2 ex_i + \alpha_3 ex_i^2 + \alpha_4 mt_i + \alpha_5 que_i + \alpha_6 sk_i + \alpha_7 al_i + \beta_9 bc_i + \alpha_9 sex_i + \alpha_{10} otl_i + \alpha_{11} en_i + \alpha_{12} otm_i + \beta_{13} s_i \] (4.107)

where \( ed_i \) is years of education\(^7\), \( ex_i \) is years of work experience\(^8\), \( qu_i \) is a dummy variable for individuals living in Quebec, \( mt_i \) is for the maritimes, \( sk_i \) denotes Manitoba and Saskatchewan, \( al_i \) is Alberta, \( bc_i \) is BC, \( sex_i \) is 1 if person \( i \) is male, \( fr_i \) and \( otl_i \) are dummies for french and non-french and non-english as first languages, \( otm_i \) and \( s_i \) denote single and non-married and non-single respectively.

The next step in replicating Taubman’s technique is to construct a new vector \( y_2 \) by the following

\[ y_2 = y - X_2 \beta_2 \] (4.108)

where the matrix \( X_2 \) contains all of the explanatory variables related to education, including the interaction with age. The vector \( y_2 \) is then interpreted as the amount of earnings that is not a return to education. Then I compare the standard deviation of \( y_2 \) to the standard deviation of earnings \( y \). The difference between these two standard deviations is assumed to be the contribution of education to earnings inequality.

The results show that the standard deviation of \( y \) is 16780 while the standard deviation of \( y_2 \) is 15849. Thus by this measure the returns to education increase inequality by 6%. Compare this with the increase of 4% reported by Taubman.

The technique used by Layard and Zabalza is similar to that of Taubman. Layard

\(^7\)See data appendix
\(^8\)See data appendix
and Zabalza's technique involves estimating the following model

\[ f = X\alpha + \nu \]  

(4.109)

where \( f = y/\mu \) and \( \mu = \sum_{i=1}^{N} y_i/N \) is the arithmetic mean of the earnings distribution and \( X \) is a vector of personal characteristics. Initially it contains only education and experience, with no interaction. That is

\[ f_i = \alpha_0 + \alpha_1 ed_i + \alpha_2 age_i + \nu_i \]  

(4.110)

where \( ed_i \) is number of years of education for person \( i \) and \( age_i \) is the age of person \( i \).

The next step is to calculate the portion of \( f \) generated by the education variables. This is denoted by \( X_2\alpha_2 \). The final step in the Layard and Zabalza analysis is to compare the variance of \( X_2\alpha_2 \) with the variance of \( f \). The variance of \( X_2\alpha_2 \) is considered the contribution of education to measured inequality. The results of my analysis show that the squared coefficient of variation of \( X_2\alpha_2 \) is 0.0341 while the index for \( f \) is 1.0022. Thus education is said to contribute 3.4\% of the earnings inequality. Compare this to the analysis of Layard and Zabalza which indicates that education contributes about 20\%.

I repeat the analysis with a much more general specification of the earnings equation.

The specification is given by

\[ f_i = \beta_1 + \beta_2 ed_i + \beta_3 ed_i age_i + \beta_4 age_i + \beta_5 age_i^2 + \beta_6 mt_i + \beta_7 que_i + \beta_8 sk_i + \beta_9 ab_i + \beta_{10} bc_i + \beta_{11} sex_i + \beta_{12} fr_i + \beta_{13} otli_i + \beta_{14} omtn_i + \beta_{15} s_i + \chi_i \]  

(4.111)

In equation 4.111 the variables are \( ed_i \) is estimated number of years of schooling\(^9\), \( age_i \) is the age of person \( i \), \( qu_i \) is a dummy variable for individuals living in Quebec, \( mt_i \) is for the maritimes, \( sk_i \) denotes Manitoba and Saskatchewan, \( ab_i \) is Alberta, \( bc_i \) is BC, \( sex_i \)

\(^9\text{See data appendix}\)
is 1 if person $i$ is male, $fr_i$ and $otl_i$ are dummies for french and non-french and non-english as first languages, $otm_i$ and $s_i$ denote single and non-married and non-single respectively. The squared coefficient of variation of $f_i$ is again 1.0022, with the education variables, including the interaction, contributing .1746. The results therefore have changed substantially when a more general earnings specification is used with education now contributing 17%. The likely reason for the drastic increase is the inclusion in this model of an interaction between education and age. Since age varies by 45 years over the sample, this will obviously increase the variability of the returns to education calculated in Layard and Zabalza's way.

The analysis done by Taubman and the analysis done by Layard and Zabalza are essentially the same, except in the measure that they use for the contribution to inequality. Taubman defines the contribution of education as

$$C = \text{Var}(y) - \text{Var}(y_2) = \text{Var}(y - y_2) + 2\text{Cov}(y, y - y_2)$$  \hfill (4.112)

while Layard and Zabalza define it as

$$C = \text{Var}(X_2\alpha_2)$$  \hfill (4.113)

Taubman attributes all of the covariance between the return to education and the return to other factors, to education while Layard and Zabalza attribute none of this covariance to education.

The preceding discussion highlights some problems with these empirical strategies. The theory behind the decomposition of the inequality measure used in the analysis has not been adequately explained. Without this explanation it is not obvious whether the decomposition methods used by these authors is appropriate for the questions they address. As well the covariance between earnings sources or more generally any measure of the interaction should, if considered, be treated as a source of inequality distinct
from the individual earnings sources and not simply arbitrarily assigned to one source of earnings or another. Taubman obviously does not do this, Layard and Zabalza have a better approach that is very similar to the approach I use in the later sections.

The results of the two papers cited, along with the replication of their analysis presented here indicate several things. Firstly, education does affect inequality but there is disagreement about how much it affects it. From 4% from Taubman’s analysis to 20% for Layard and Zabalza’s measure. This disagreement is partially a result of the different ways of measuring the effect of different income sources on inequality indices. Taubman’s measure of the contribution of education includes all of the covariance between returns to education and other returns while Layard and Zabalza do not attribute this covariance to education. The quantitative differences between my analysis and Layard and Zabalza’s is likely because they use family income instead of individual earnings and also consider the education level of both spouses in their measure of education. Both of these differences would increase the amount of inequality explained by education.

The first paper in this thesis is a step in the direction of standardizing the technique for estimating the contribution of an income source to income inequality. Also, from the quantitative differences obtained when moving to a more general specification of the earnings equation with the Layard and Zabalza technique, it seems that the specification of the earnings equation has an important effect. Subsequent sections therefore address these issues.

4.5 Returns to Education

In this section I describe the method I use to decompose labour earnings into two factor components, a part due to differences in education and a part due to other factors. The procedure I adopt is to first estimate an earnings equation and then to use the estimated
earnings equation to remove the effect of differing education levels.

The estimation of human capital earnings equations has a long history in labour economics, including work by Mincer (1974) and Welch (1970). These models typically specify a potential earnings equation for person $i$ of the general form

$$y_i^* = f(a_i, x_i) + \epsilon_i$$

(4.114)

where $y_i^*$ is the potential earnings of person $i$, $a_i$ is the education level of person $i$, $x_i$ is a vector of personal characteristics for person $i$ and $\epsilon_i$ is a random disturbance term. The parameters of the function $f$ are to be estimated.

A problem arises in estimating equation 4.114. The true values of $y_i^*$ are observed only when $y_i^*$ is positive, otherwise the observed value of $y_i$ is zero. This results in the spike in the distribution that I discussed in section two. If I ignore this feature of the data and just run the regression on the observed $y_i$s, the resulting estimates of the parameters $a_i$ will be biased.

Various methods of dealing with this problem exist in the literature. The method I choose to deal with the problem is the Tobit method proposed by Tobin (1958). This technique makes the method of dealing with the spike in the decomposition of earnings straightforward and intuitive.

The Tobit method involves constructing a likelihood function for observed $y_i$ which includes both the probability of being in the spike and having zero earnings, and the probability density function for the positive earnings. More specifically note that observed $y_i$ will be zero from equation 4.114 only if $\epsilon_i \leq -f(a_i, x_i)$. The probability of this happening, given normality of the $\epsilon_i$ is $\Phi(-f(a_i, x_i)/\sigma)$ where $\sigma$ is the standard deviation of the error and $\Phi$ is the standard normal distribution function. If $y_i$ is not zero then

\[\text{See Willis (1986) for a discussion of estimating earnings equations.}\]
\[\text{See also Greene (1990)}\]
\[\text{See section six.}\]
Chapter 4. The Contribution of Education to Earnings Inequality

the error follows a normal distribution with density \( \phi((y_i - f(a_i, x_i))/\sigma) \). The density function for a given observation is then given by

\[
\phi \left( \frac{-f(a_i, x_i)}{\sigma} \right)^{I(y_i \leq 0)} \phi \left( \frac{y_i - f(a_i, x_i)}{\sigma} \right)^{1-I(y_i \leq 0)}
\]

(4.115)

where \( I(y_i \leq 0) \) is an index function that is 1 if \( y_i \leq 0 \) and 0 otherwise. The above density is used to construct a likelihood function and the parameters of the function \( f \) are estimated by maximum likelihood.

The Survey of Family Expenditures does not contain direct observations on the level of education, instead the data has only categories of education level attained. Thus I estimate a version of the earnings equation where the effect of education is captured by the use of dummy variables. Conspicuous in their absence for the earnings regression are variables such as industry, occupation, and union status. The reason that I do not include these variables, even though they have often been shown to be significant determinants of earnings, is that I want my measure of returns to education to include, as a potential return to education, the ability to work in some jobs that may be higher paying.

Consider occupation, as part of my construction of the counterfactual incomes, I predict the earnings distribution that would result if everyone had the same education level. The presence of occupation variables would confuse the issue of what is a return to education. Suppose we did live in a world where everyone had the same level of formal education. Then formal education would not qualify anyone to work as a medical doctor. Nevertheless it is likely that some sort of informal training in medicine would arise. Thus the return to working in the medical profession would include a return to this informal training. It would be impossible to separate the training effect from the occupation effect in the construction of the counterfactual earnings levels. Even allowing for different education levels it is still likely that the measured return to occupation has a training effect.
Due to the difficulty in separating the legitimate occupational effect on earnings from the education effect, I leave information on occupation out of the estimation completely. The estimated coefficients on education are unconditional on being in a particular occupation or industry. Thus my measure of returns to education includes such things as the ability to work in some industries that may pay higher or which are more likely to be unionized, but that also may require a greater level of training. The bias introduced by not being able to separate the training effect from the occupation effect is a potential limitation of the analysis.

The tobit regression of the earnings equation \( f \) is based on the following earnings equation

\[
y_i = \alpha_1 + \alpha_2 d_i + \alpha_3 d_3 + \alpha_4 d_4 + \alpha_5 d_5 + \alpha_6 d_2a_i + \alpha_7 d_3a_i + \alpha_8 d_4a_i \\
+ \alpha_9 d_5a_i + \alpha_{10} m_2 + \alpha_{11} m_3 + \alpha_{12} m_4 + \alpha_{13} m_5 + \alpha_{14} m_2a_i + \alpha_{15} m_3a_i \\
+ \alpha_{16} m_4a_i + \alpha_{17} m_5a_i + \alpha_{18} a_i + \alpha_{19} a_i^2 + \alpha_{20} m_i + \alpha_{21} q_i. \quad (4.116)
\]

Where \( d_i \) is a dummy variable that indicates that person \( i \) is in education category \( l \), \( a_i \) is the person's age, \( q_i \) is a dummy variable for individuals living in Quebec, \( m_t \) is for the maritimes, \( sk_i \) denotes Manitoba and Saskatchewan, \( ab_i \) is Alberta, \( bc_i \) is BC, \( sex_i \) is 1 if person \( i \) is male, \( fr_i \) and \( ot_i \) are dummies for french and non-french and non-english as first languages, \( om_t \) and \( s_i \) denote single and non married and non single respectively. \( d_l a_i \) is the product of \( d_l \) and \( a_i \), while variables beginning with \( m \) are ones which are the product of the gender variable and the indicated variable. There is thus 5 separate categorical variables with separate dummies for each. To avoid the dummy variable trap each categorical variable should have a reasonable number of people omitted. The

\(^{13}\)Lemieux (1993) provides some evidence for Canada that there are definite patterns to unionization with respect to skill level.
Chapter 4. The Contribution of Education to Earnings Inequality

proportion of people omitted goes from approximately 14% for the education status to approximately 82% for the marital status dummy. Overall there are over 200 individuals who are in the base category of female, married, english speakers, living in Ontario, with less than nine years of schooling.

The tobit estimates of equation 4.116, for the data set already described, are given in column three of Table 4.2 with the asymptotic standard errors in column four. Note that these are the estimates of the parameters divided by the estimated standard deviation of \( x \). Thus for example the reported coefficient for \( y \) is the reciprocal of the standard error of the estimate. The OLS estimates of equation 4.116 are in columns one and two. Inspection of the tobit estimates shows that education increases the expected earnings and that this effect increases with age. Of the rest, most of the explanatory variables decrease expected earnings, with the exception of being male and single, from the base case of a female, living in the Ontario, with English as a first language, and with a marriage status that is married. Both language variables reduce earnings as well as being neither married nor single. Several of the variables, notably \( d2age, d3age, d5age, al, fr atm, \) and \( s \) are insignificantly different from 0 at a size of .05. Examining the interaction between the education variables and the gender dummy it can be seen that men have a returns to education that starts lower but increases at a faster rate with age. The results of this regression are consistent with previous studies, the return to education is positive and earnings are increasing in experience. The results are essentially the same for the OLS estimates except for the interaction of education categories three and four which are insignificantly different from zero. The estimated values for both estimation techniques appear to be in accord with previous Canadian studies of wage and earnings determination, for example Green (1991), Simpson (1985) and Kumar and Coates (1982).

Since the focus of this study is on the contribution of education to income inequality special attention should be taken with these parameters. The above specification of
<table>
<thead>
<tr>
<th>Parameter</th>
<th>OLS Value</th>
<th>OLS Standard Error</th>
<th>Tobit Parameter</th>
<th>Tobit Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-35089</td>
<td>2208</td>
<td>-3.0980</td>
<td>.18315</td>
</tr>
<tr>
<td>d2</td>
<td>9275.5</td>
<td>1592</td>
<td>.7440</td>
<td>.1363</td>
</tr>
<tr>
<td>d3</td>
<td>12166</td>
<td>2054</td>
<td>.9510</td>
<td>.1710</td>
</tr>
<tr>
<td>d4</td>
<td>18152</td>
<td>1987</td>
<td>1.3590</td>
<td>.1646</td>
</tr>
<tr>
<td>d5</td>
<td>16333</td>
<td>2498</td>
<td>1.1395</td>
<td>.2025</td>
</tr>
<tr>
<td>d2age</td>
<td>-127.03</td>
<td>32.68</td>
<td>-.0051</td>
<td>.0029</td>
</tr>
<tr>
<td>d3age</td>
<td>-113.21</td>
<td>47.31</td>
<td>-.0025</td>
<td>.0040</td>
</tr>
<tr>
<td>d4age</td>
<td>-207.57</td>
<td>45.39</td>
<td>-.0079</td>
<td>.0038</td>
</tr>
<tr>
<td>d5age</td>
<td>-30.220</td>
<td>59.15</td>
<td>.0065</td>
<td>.0048</td>
</tr>
<tr>
<td>md2</td>
<td>-108.17</td>
<td>1304</td>
<td>-.5687</td>
<td>.1068</td>
</tr>
<tr>
<td>md3</td>
<td>-6639.7</td>
<td>2317</td>
<td>-1.0461</td>
<td>.1843</td>
</tr>
<tr>
<td>md4</td>
<td>-10731</td>
<td>2294</td>
<td>-1.3432</td>
<td>.1809</td>
</tr>
<tr>
<td>md5</td>
<td>-8053</td>
<td>2854</td>
<td>-1.1333</td>
<td>.2239</td>
</tr>
<tr>
<td>md2age</td>
<td>101.60</td>
<td>27.48</td>
<td>.0130</td>
<td>.0022</td>
</tr>
<tr>
<td>md3age</td>
<td>260.85</td>
<td>57.73</td>
<td>.0222</td>
<td>.0046</td>
</tr>
<tr>
<td>md4age</td>
<td>403.77</td>
<td>56.61</td>
<td>.0311</td>
<td>.0045</td>
</tr>
<tr>
<td>md5age</td>
<td>387.88</td>
<td>69.33</td>
<td>.0282</td>
<td>.0054</td>
</tr>
<tr>
<td>age</td>
<td>2092.6</td>
<td>83.33</td>
<td>.1645</td>
<td>.0069</td>
</tr>
<tr>
<td>age²</td>
<td>-24.050</td>
<td>.8510</td>
<td>-.0021</td>
<td>.00007</td>
</tr>
<tr>
<td>mt</td>
<td>-4939.9</td>
<td>355.4</td>
<td>-.3885</td>
<td>.0285</td>
</tr>
<tr>
<td>que</td>
<td>-3035.3</td>
<td>485.2</td>
<td>-.2820</td>
<td>.0390</td>
</tr>
<tr>
<td>sk</td>
<td>-2430.9</td>
<td>399.8</td>
<td>-.1722</td>
<td>.0318</td>
</tr>
<tr>
<td>al</td>
<td>-638.08</td>
<td>418.5</td>
<td>-.0541</td>
<td>.0330</td>
</tr>
<tr>
<td>bc</td>
<td>-2910.8</td>
<td>419.4</td>
<td>-.2366</td>
<td>.0335</td>
</tr>
<tr>
<td>sex</td>
<td>11258</td>
<td>613.2</td>
<td>1.2520</td>
<td>.0537</td>
</tr>
<tr>
<td>otl</td>
<td>-2194.8</td>
<td>373.7</td>
<td>-.1256</td>
<td>.0301</td>
</tr>
<tr>
<td>fr</td>
<td>-358.86</td>
<td>448.6</td>
<td>-.0106</td>
<td>.0361</td>
</tr>
<tr>
<td>otm</td>
<td>-4.7781</td>
<td>400.5</td>
<td>-.0118</td>
<td>.0331</td>
</tr>
<tr>
<td>s</td>
<td>35.889</td>
<td>422</td>
<td>.0425</td>
<td>.0332</td>
</tr>
<tr>
<td>y = 1/σ</td>
<td>-</td>
<td>-</td>
<td>.00006</td>
<td>.0000005</td>
</tr>
<tr>
<td>R²</td>
<td>.3957</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEE</td>
<td>13070</td>
<td>15951</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LLF</td>
<td>-140873</td>
<td>-110770.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>12929</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Estimates of the Earnings Equation.
Chapter 4. The Contribution of Education to Earnings Inequality

the earnings equation allows the return to education to be different for the two most important demographic groups. This difference is allowed to affect both the level of the returns to education and the way that the return to education changes with the experience of the individual. The results from the tobit regression indicate that the difference does seem to be important. Men seem to have a lower initial return to education but a faster rate of increase over time. This is a surprising result, seeming to run counter to conventional wisdom. Part of the explanation may lie with the gender dummy. Being male tends to increase earnings, irrespective of education, by at least the amount that men have a decrease in the return to education.

Given the importance of the return to education for the subsequent analysis a hypothesis test is performed to test the hypothesis that the returns to education for men and women are the same. That is, I am testing the null hypothesis that $\alpha_{10}$ through $\alpha_{17}$ in equation 4.116 are equal to zero. In appendix two I provide the results for a regression where the restriction is imposed. $-2(LLF_U - LLF_R)$ is approximately distributed $\chi^2$ with 8 degrees of freedom. The value of this statistic is 146.4 and the critical value of the $\chi^2$ distribution at size .05 is 15.507 so the null is rejected, in fact it is rejected at all reasonable sizes, and I conclude that the structural difference in returns to education between men and women is likely to be important.

Given the presence of some coefficients that are insignificantly different from zero it may be questioned whether these variables should be used in the generation of the counterfactual earnings. The question is really whether the variables with insignificant coefficients should be dropped and the model reestimated. Simply dropping the insignificant coefficients and using the remaining estimates is clearly wrong since the model being used to predict would not be the one that was used to estimate the parameters.

Is it appropriate to simply drop variables when they are insignificant? The answer, both economically and econometrically is no. If the economic theory suggests that a
given variable be included in a regression, then to leave it out is to subject the model to possible misspecification bias. Sometimes this is necessary, such as when the required data is not available, but as general rule it is something that should be avoided if possible. The fact that a coefficient is insignificantly different from zero does not mean that the value of the parameter in the true model is zero. The insignificance merely means that the variance of the estimated parameter is large relative to the estimated value. In the absence of other reasons for the variable to be excluded it would be incorrect to exclude on the basis of low t-statistics.

While the economics provides the most compelling argument for not omitting variables with low t-statistics there is another, statistical, argument. If all variables with t-ratios less than a given number in absolute value are dropped, then the remaining estimates no longer have the standard t-distribution. They have a censored distribution that is related to the t. If this is the case then the meaning of significant t-ratios is unclear.

Given these two arguments for using all of the variables, regardless of significance level I proceed with the construction of the decomposition of earnings using the regression results from the earnings equations. Denote person i’s kth counterfactual earnings level to be \( \bar{y}_i^k \), i.e. \( \bar{y}^k \) is the earnings distribution when everyone who’s actual level of education is at or below level \( k \) is given a counterfactual level of education equal to their actual level. Everyone else has a counterfactual education level of \( k \). \( \bar{y}_i^k \) is constructed by first generating the \( k \)th counterfactual education dummy variable, \( \bar{d}_i^k \), in the following way. If \( d_l^i = 1 \) for some \( l \geq k \) then \( \bar{d}_i^k = 1 \) and \( \bar{d}_i^l = 0 \) for \( l \neq k \). If \( d_l^i = 0 \) for all \( l \geq k \) then \( \bar{d}_i^k = d_l^i \) for all \( l = 1, \ldots, 5 \). The predicted counterfactual earnings \( \bar{y}_i^k \) is generated from the estimated earnings equation, using the education variables \( \bar{d}_i^k \) instead of the actual education variables, and including the estimated error term in the following way; denote by \( \bar{y}^k \) the vector of base earnings with the \( k \)th education level as the base education level. For the initial analysis I ignore the effect that education has on the probability of being
employed. Thus the counterfactual distribution for the initial empirical analysis assumes that an individual who is employed with the actual education level will also be employed with the counterfactual education level. The vector of final counterfactual earnings is generated by equation 4.117

\[
\hat{y}_i^k = y_i - \hat{\alpha}_2d2 - \hat{\alpha}_3d3 - \hat{\alpha}_4d4 - \hat{\alpha}_5d5 - \hat{\alpha}_6d2age - \hat{\alpha}_7d3age - \\
\hat{\alpha}_8d4age - \hat{\alpha}_9d5age - \hat{\alpha}_{10}md2 - \hat{\alpha}_{11}md3 - \hat{\alpha}_{12}md4 - \hat{\alpha}_{13}md5 - \hat{\alpha}_{14}md2age - \\
\hat{\alpha}_{15}md3age - \hat{\alpha}_{16}md4age - \hat{\alpha}_{17}md5age + \hat{\alpha}_2cf2 + \hat{\alpha}_3cf3 + \hat{\alpha}_4cf4 + \hat{\alpha}_5cf5 + \\
\hat{\alpha}_6cf2age + \hat{\alpha}_7cf3age + \\
\hat{\alpha}_8dcf4age + \hat{\alpha}_9cf5age + \hat{\alpha}_{10}md2 + \hat{\alpha}_{11}mcf3 + \\
\hat{\alpha}_{12}mcf4 + \hat{\alpha}_{13}mcf5 + \hat{\alpha}_{14}mcf2age + \hat{\alpha}_{15}mcf3age + \hat{\alpha}_{16}mcf4age + \hat{\alpha}_{17}mcf5age. \quad (4.117)
\]

In equation 4.117 the ‘hat’ denotes the estimates of the parameters for equation 4.116. In subsection 3.6.1 I use the OLS estimates from table 4.2 to construct the earnings, in subsequent sections, 3.6.2 and later, I use the tobit estimates to construct the counterfactual earning so the values from table 4.2 must be multiplied by the standard error of the estimate (SEE from the table) to get the appropriate value of the parameters for use in constructing the counterfactual earnings.

While equation 4.117 may look intimidating all it really does is remove the effect of actual education from earnings, to leave the earnings that cannot be attributed to education, and then add back the predicted effect of the counterfactual education levels. Note that in the case where the counterfactual education level is less than nine years of schooling, this reduces to the earnings decomposition method of Taubman (1975). I now define the vector of earnings due to differences in education to be \(y^{ek} = y - \hat{y}^k\). The inequality analysis in the next section is undertaken using \(y^{e1}\) and \(\hat{y}^1\). In other words the base level of education is category 1, less than nine years formal training.
Chapter 4. The Contribution of Education to Earnings Inequality

The preceding method of constructing the counterfactual distribution does not allow education to affect the probability of employment. I now outline a method of construction of the counterfactual earnings which does allow for this effect. Consider the sample distribution of earnings. This distribution has two important parts. The first is the regular distribution of earnings for those who are employed. The second is the spike of individuals who have earnings of zero because they are not working. In the sample of 12,929 people, 3,167 are not working and 9,762 are employed. I want a decomposition of earnings that takes into account the change in the probability of being in the spike at zero earnings when the education level changes. The use of the tobit model allows me to include the fact that one of the more important elements of the return to education is that increased education can increase the chances of an individual being employed and earning a positive income instead of being unemployed and having an income of zero.

To take this effect into account the construction of the counterfactual income level is changed slightly from that outlined previously. When the vector of counterfactual earnings is calculated from equation 4.117, it is regarded as potential earnings. If these potential counterfactual earnings are negative the individual is assumed unemployed at the counterfactual education level and receives $\bar{y}_i^k$ of zero. If, however, the potential counterfactual earnings are positive then the individual is assumed to be employed at the counterfactual education level and receives income $\tilde{y}_i^k$. The return to education for individual $i$ is then determined by $y_i^e = y_i - \bar{y}_i^1$.

One initially appealing method of determining the contribution of education to inequality is to use the $R^2$ of the regression of $y$ on the personal variables, including education, with the $R^2$ of the same regression excluding education variables. The reason that this is not appropriate is that, if any of the personal characteristics, such as marital status, are correlated with the education variables, then some of the variation due to education will be explained by the personal characteristics in the restricted model and
the resultant $R^2$ will not be true measure of the amount of variation due to education. Thus the appropriate method to get a measure of the base earnings is to estimate the full model and then remove the effect of the education variables.

It is important to recognize just what the respective counterfactual earnings contain. In generating these base level distributions, only partial equilibrium effects are being captured. The same is true of any external effects of education. To the extent that these are important, the subsequent analysis is limited\(^{14}\).

### 4.6 The Effect of Education on Inequality.

I now use the counterfactual earnings distributions generated in the previous section to examine how the contribution of education to earnings affects overall earnings inequality. As a first step at understanding the contribution of education to earnings inequality examine figure 4. This is a graph of the well known Lorenz curve for both the predicted earnings distribution and the counterfactual distribution of the base level of education using the estimated earnings equation from the previous section\(^{15}\). The Lorenz curve shows the percentage of population, ranked from the poorest to richest, on the horizontal axis, and the percentage of aggregate earnings earned by that segment of the population on the vertical axis. The Lorenz curve for the actual earnings distribution is higher at all percentiles of the population than the Lorenz curve corresponding to the vector of counterfactual earnings. Thus the actual earnings distribution is unambiguously more equally distributed, than the counterfactual distribution, however from the Lorenz curve this difference appears to be relatively small.

Although the Lorenz curve is useful as a visual aid to understanding inequality it does not provide all of the information desired. For example no scale effects are captured by a

---

\(^{14}\)This question is dealt with in much more detail in Chapter 3.

\(^{15}\)The counterfactual in this case includes the effect of education on employment.
Lorenz curve. Inequality in an earnings source that is a very small part of total earnings is treated in the same way as an earnings source that is a very important part of total earnings. Decomposing an inequality index provides more information about how the earnings sources $\bar{y}^1$ and $y^e$ contribute to overall inequality because it provides information about both the scale of the source of earnings in relation to the total earnings as well as the interaction between various sources of earnings. At this stage I will concentrate attention on the Atkinson–Kolm–Sen (AKS) inequality index as this is the more widely known of the two popular ethical indices.

Assume that society has preferences over alternative earnings distributions which can be represented by the function $w = w(y)$ where $w$ is $S$-concave, increasing and every associated social indifference curve crosses the ray of equality defined by the equation $y = a1^N$, where $1^N$ is a $N$-vector of ones. The equally distributed equivalent (EDE) earnings function $\xi = \Xi(y)$ is defined implicitly by the equation $w(y) = w(\xi 1^N)$. The EDE is the amount of earnings which, if distributed equally, would give the same level of welfare as the current earnings distribution and is an exact representation of society's preferences over earnings distributions. The AKS inequality index is defined as $I(y) = 1 - \Xi(y)/\mu(y)$ where $\mu(y)$ is the arithmetic mean of the earnings distribution. The AKS index gives a measure of the percentage of current total earnings that could be given up by the society and still obtain the same level of welfare, providing that the remainder of the earnings is split evenly\footnote{See Blackorby and Donaldson (1978) or Chakravarty (1990).}. I then decompose this inequality index by factor components as described in the previous paper. There are two components to earnings $\bar{y}^1$, the base level of earnings, and $y^e$ the earnings due to education. Thus the decomposition takes the form

$$I(y) = s^1 I(\bar{y}^1) + s^e I(y^e) + s^1 [I(y) - I(\bar{y}^1)] + s^e [I(y) - I(y^e)]$$ (4.118)
Chapter 4. *The Contribution of Education to Earnings Inequality*

where \( s^j = \mu^j / \mu \), the ratio of the mean of income source \( j \) to the mean of total income, is the share of aggregate income going to source \( j \), \( I(\tilde{y}^1) \), \( I(y^e) \), and \( I(y) \) are the AKS inequality indices for the counterfactual level of income, the returns to education, and aggregate income respectively. The first step in the construction of ethical inequality indices is the selection of an EDE function. The S-Gini EDE function was introduced by Donaldson and Weymark (1980) and axiomatized by Bossert (1990). The specific functional form of the S-Gini EDE function is

\[
\Xi(y) = \frac{\sum_{i=1}^{N}\left[i^\delta - (i - 1)^\delta\right]\tilde{y}_i}{N^\delta} \quad (4.119)
\]

In equation 4.119, \( \tilde{y} \) is the 'welfare ranked' permutation of the earnings vector \( y \) such that \( \tilde{y}_i \geq \tilde{y}_{i-1} \) for all \( i = 1, ..., N \), and \( \delta \) is an inequality aversion parameter. If \( \delta = 1 \) then the S-Gini SEF becomes a linear SEF and \( \Xi(y) = \mu \). This means that income contributes the same to social welfare no matter who has it. The social indifference curves are planes in the space of incomes. In the case of two individuals, as in figure 5, the social indifference curves are straight lines. If \( \delta = 2 \) then equation 4.119 is the standard Gini SEF as defined by the AKS inequality index equal to the ratio of the area between the Lorenz curve and the line of perfect equality out of the origin, and the total area beneath the ray of perfect equality. A representative social indifference curve for this case is also given in figure 5. As \( \delta \) tends to \( \infty \), the S-Gini SEF approaches the maximin SEF and the AKS inequality index becomes one minus the ratio of the lowest income to the mean income.

The reason for the S-Gini is mostly pragmatic. The data has many observations with zero earnings, or a zero contribution of education. Given this I need an SEF which can handle zeros. The S-Gini can handle zeros, most other common SEF's such as the mean of order \( r \) cannot.
Table 4.3: Decomposition of $I(y)$ for Earnings Decomposition 2

### 4.6.1 Initial Estimation

Table 4.3 presents the decomposition of the S-Gini AKS index for the earnings decomposition described in the previous section by equation 4.117. It uses the estimated parameters for the straightforward OLS regressions, based on equation 4.116 but without the tobit correction. This is done for the whole sample, the sample restricted to those between 30 and 40 years of age, and those between 50 and 60 years of age. This separation into subsamples is done for two reasons. The first being to see if the effect of education on inequality varies with age. The second is that there is substantial reason to believe that there are life-cycle inequality effects arising from life-cycle earnings patterns. Therefore it may not be as bad if an older person earns more than a younger person than if two similarly aged individuals have different earnings. The indices are calculated using values of 1, 2 and 5 for the inequality aversion parameter $\delta$. These values provide a fairly wide range of inequality aversion. To illustrate this point, a graphical depiction of an iso-welfare line for the three functions in the case when $N = 2$ is shown in figure 5. Column 4 of table 4.3 shows that, for the earnings equation, 37% of the total earnings in the entire sample is a return to education above the base level. This value remains
Chapter 4. The Contribution of Education to Earnings Inequality

constant for the subsample aged 30-40 and falls to 28% for the cohort aged 50-60.

Examining the inequality in aggregate earnings in column 6, it is apparent that the inequality in the sample of 30-40 year olds is substantially less than the inequality in earnings for both the other subsample and the complete sample. This is probably a result of the lower experience of these workers in relation to the older cohort. Since experience interacts with the other variables, especially education, a lower experience level will likely produce a more homogenous earnings distribution. Compare this with the life-cycle pattern reported by Jenkins (1992) for the UK, where he finds that inequality is slightly less for the older cohorts. The measured inequality for the S-Gini when the inequality aversion parameter is $\delta = 5$ is, as expected, much higher for all of the subsamples.

Turning to the decompositions, as shown by columns 1 and 2, the returns to the base level of education are more unequally distributed than the returns to education in all cases. It seems that returns to education are less unequally distributed in the age group 30-40 than they are in the other two samples. This may reflect the reduced interaction effect in the wage equation between education and experience. At the same time the base level of earnings is also more equally distributed for the 30-40 subsample than for the whole sample or the 50-60 year olds which are essentially the same.

The interaction term in the decomposition, column 5 is negative in all cases, indicating that the inequality in the base earnings and the inequality in the returns to education partially cancel each other out. For example the rank of person $i$ in the distribution of returns to education has a low correlation with his or her rank in the base returns. Anything other than perfect positive rank correlation will result in some cancellation of inequality by the two sources. The actual sign of the interaction is not a surprise since it can be shown that for the S-Gini inequality index the interaction term will always be

17The discrepancy is likely a result of Jenkins' use of family disposable income instead of earnings.
non-positive. The extent of this interaction seems quite significant, when $\delta = 2$ having the effect of an increase of average earnings of 5% for the whole sample, slightly lower for the 30-40 year olds and the 50-60 year olds.

Table 4.4 gives the total contribution of the base returns, the returns to education, and the interaction term to total measured inequality. Table 4.4 can be interpreted in the following way. The AKS inequality index gives the percentage of aggregate income that could be removed while maintaining the same level of welfare, providing that the remaining income is distributed equally. For the full sample, when $\delta = 2$ inequality in earnings wastes 53% of total income. Column two in table 4.4 indicates that of that 53, 20 is attributable to inequality in returns to education. Thus inequality in the return to education causes a loss of welfare equivalent to a decrease in aggregate income of 20%, provided that the remainder is distributed equally.

Examining the cohort effects it is apparent that when $\delta = 2$ the lowest contribution of education is for the 30-40 year old cohort. This cohort also has the lowest contribution of the base level of education. Their value for education is 17 compared to 20 for the full sample. The 30-40’s contribution of the base level is 34 when $\delta = 5$ compared to 39 for the full sample. Things are slightly different when $\delta = 5$. In this case the 50-60 cohort

<table>
<thead>
<tr>
<th>Age</th>
<th>$\delta$</th>
<th>$s^1I(\bar{y}^1)$</th>
<th>$s^eI(y^e)$</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.3865</td>
<td>0.1974</td>
<td>-0.0550</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.6020</td>
<td>0.3336</td>
<td>-0.0361</td>
</tr>
<tr>
<td>30-40</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.3411</td>
<td>0.1731</td>
<td>-0.0477</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.5664</td>
<td>0.3190</td>
<td>-0.0386</td>
</tr>
<tr>
<td>50-60</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.4673</td>
<td>0.1965</td>
<td>-0.0463</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.8283</td>
<td>0.2477</td>
<td>-0.2292</td>
</tr>
</tbody>
</table>

Table 4.4: Amount of Measured Inequality Attributable to Different Sources
has the lowest contribution of education but it also has the highest contribution of the base level. The contribution of education is approximately the same for the 30-40 cohort and the full sample at 32 and 33 respectively. The two samples are also similar in the effect of the base level at 60 for the full sample and 57 for the 30-40 cohort.

An advantage of using the decomposition of the AKS inequality indices that I use is that the terms of the decomposition inherit the numerical significance of the AKS index. This means that it makes sense to speak of education contributing a specific percentage to measured earnings inequality. This feature provides a nice way of examining whether the differences outlined above are just a result of the different measured inequality or are actually a result of systematic differences in the effect of the individual components. It also provides a way of comparing the analysis here with that of Taubman, and Layard and Zabalza. Table 4.5 therefore presents a percentile representation of the results for the decomposition of earnings.

In column 2 of table 4.5 it can be seen that for the overall sample, the direct effect of education is to increase measured inequality by approximately 37% when \( \delta = 2 \) and by 37% when \( \delta = 5 \). Another way of looking at this is that 37% of total measured inequality is contributed by the returns to education. Compare this with the results of Taubman
who found that 4% of earnings inequality were a result of returns to education or Layard and Zabalza who found 20% was a result of returns to education or to my replication of their analysis which found 6% and 3.4% respectively. At the same time the contribution of the base level of earnings is 73% and 67% respectively as shown by column 1. Thus the interaction term for the full sample actually has a fairly large effect on the inequality in the full sample reducing inequality by about 10% or 4%.

The results for the sample aged 30-40 are similar, the direct effect of education contributes 37% of measured inequality when $\delta = 2$, rising slightly to 38% when $\delta = 5$. The interaction effect and the contribution of the base level of earnings remain approximately the same as in the full sample. The sample of people aged 50-60 shows a similar pattern of effects as the other two samples. Returns to education contribute 28% of the measured inequality in this case. When $\delta = 5$ the results show that the return to education contributes approximately the same to overall inequality. The interaction effect is less in this sample than in the other two, only reducing inequality by 4% when $\delta = 2$ and 1% when $\delta = 5$.

I now briefly consider the results from the marginal interactive decomposition for the straightforward OLS earnings model. The marginal interactive decomposition is given by the following equation,

$$ I(y) = \sum_{j=1}^{J} s^j (I(y) - I(y^{-j})) + C^I(y^{-1}, ..., y^{-J}, y). \quad (4.120) $$

The first $J$ terms in the interactive decomposition in equation 4.120 give the marginal contribution of source $j$ to overall inequality while the last term is an interaction term which provides a reference level of inequality. The results for the OLS earnings model are given in table 4.6

In column two of table 4.6 it shows that the marginal effect of the return to education is to decrease measured inequality, equivalent to an increase in EDE of 3% for the full
Chapter 4. The Contribution of Education to Earnings Inequality

<table>
<thead>
<tr>
<th>Age</th>
<th>δ</th>
<th>$\bar{y}^1$</th>
<th>$y^e$</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.0063</td>
<td>-0.0297</td>
<td>0.5649</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.0071</td>
<td>-0.0185</td>
<td>0.9251</td>
</tr>
<tr>
<td>30-40</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0098</td>
<td>-0.0335</td>
<td>0.4903</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0099</td>
<td>-0.0280</td>
<td>0.8649</td>
</tr>
<tr>
<td>50-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.0697</td>
<td>-0.0195</td>
<td>0.7067</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.197</td>
<td>-0.0032</td>
<td>0.9779</td>
</tr>
</tbody>
</table>

Table 4.6: Marginal Decomposition for the OLS Model

sample when $\delta = 2$. This remains fairly constant for the 30-40 and falls to 2% for the 50-60 cohort. Thus for the last subsample the equalizing effect on earnings of education is less important than it is for the sample as a whole and for the 30-40 cohort. The marginal effect of the base returns is positive in the 30-40 case indicating that this earnings source tends to increase measured earnings inequality. This marginal effect is negative for the full sample and for the 50-60 cohort.

The major results of this subsection are that education contributes about 30% of measured inequality. Education does however serve as an equalizing influence on earnings. This equalization effect is strongest for the younger 30-40 cohort, perhaps indicating the greater importance of education on earnings for this cohort relative to the other subsamples.

4.6.2 Complete Earnings Model

The analysis in the previous section used a straightforward OLS estimation of the earnings equation to decompose income. It is well known that OLS will provide biased estimates of the coefficients in the case where the data is censored. As well, no account was taken in the previous subsection of the effect of education on the probability of employment. In
Chapter 4. The Contribution of Education to Earnings Inequality

<table>
<thead>
<tr>
<th>Age</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1.00</td>
<td>0.00</td>
<td>0.5767</td>
<td>0.4233</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.6492</td>
<td>0.5228</td>
<td>0.5767</td>
<td>0.4233</td>
<td>-0.0688</td>
<td>0.5289</td>
</tr>
<tr>
<td>5</td>
<td>0.9662</td>
<td>0.9003</td>
<td>0.5767</td>
<td>0.4233</td>
<td>-0.0388</td>
<td>0.8995</td>
</tr>
<tr>
<td>30-40</td>
<td>1.00</td>
<td>0.00</td>
<td>0.5723</td>
<td>0.4277</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.5972</td>
<td>0.4394</td>
<td>0.5723</td>
<td>0.4277</td>
<td>-0.0631</td>
<td>0.4666</td>
</tr>
<tr>
<td>5</td>
<td>0.9444</td>
<td>0.8204</td>
<td>0.5723</td>
<td>0.4277</td>
<td>-0.0445</td>
<td>0.8468</td>
</tr>
<tr>
<td>50-60</td>
<td>1.00</td>
<td>0.00</td>
<td>0.6271</td>
<td>0.3729</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.6829</td>
<td>0.6660</td>
<td>0.6271</td>
<td>0.3729</td>
<td>-0.0591</td>
<td>0.6175</td>
</tr>
<tr>
<td>5</td>
<td>0.9773</td>
<td>0.9739</td>
<td>0.6271</td>
<td>0.3729</td>
<td>-0.0210</td>
<td>0.9550</td>
</tr>
</tbody>
</table>

Table 4.7: Decomposition of $I(y)$ for Earnings Decomposition

In this section I repeat the above analysis using a properly specified tobit model, outlined in the previous section, to estimate the parameters of the earnings function and use these in the decomposition of earnings which includes the employment effect, also outlined previously.

Examining column four of Table 4.7 the share of returns to education in terms of total income is slightly less than it was with the OLS estimation. The cohort aged 50-60 has the lowest proportion of their earnings in the form of returns to education and the cohort aged 30-40 and the full sample having approximately the same. The overall measured inequality in column six is identical to that previously.

Turning to the within source inequality in columns one and two it is again true that the inequality in returns to education is lower for the 30-40 cohort than for the other two subsamples. Looking at the case when $\delta = 2$ the inequality in returns to education for the cohort aged 50-60 is half again that of the 30-40 cohort, when $\delta = 5$ the difference is not so striking but the inequality in returns to education for the 50-60 cohort is still much higher than that of the 30-40 cohort. The larger inequality in returns to education for the full sample is not as dramatic but is still much higher than that of the 30-40 cohort.
Table 4.8: Amount of Measured Inequality Attributable to Different Sources

The inequality in the base levels of income is also lowest for the 30-40 cohort and highest for the 50-60 cohort although the magnitude of the difference is nowhere near that for the inequality in returns to education.

Compared to the previous OLS analysis returns to education are slightly more equally distributed but this difference is fairly minor. The interaction term in column five is again negative but is considerably smaller compared to the sample with only the OLS decomposition. It gets to approximately -.06 when \( \delta = 2 \) and is less when \( \delta = 5 \) (compare with a low of approximately -.10 with the OLS analysis). The interaction effect seems to be more important for the 30-40 cohort and the full sample than for the 50-60 cohort. This is similar to the previous OLS analysis.

Column two in table 4.8 shows that, for the full sample with \( \delta = 2 \) inequality in returns to education causes the same loss in welfare as would a 22% decrease in aggregate income if the remainder were distributed equally. This amount increases to 38% when the inequality aversion rises to 5. This is larger than the effect calculated with the OLS estimates alone. Counteracting this difference is that the inequality in the base levels is less in this case than with the straightforward OLS estimates. The inequality in the base level of education causes the same loss of welfare as would a decrease in aggregate income.
income of 38% and 56% for $\delta = 2$ and $\delta = 5$ respectively. The interaction effect between the two sources serves to increase welfare the same as a 7% increase in aggregate income when $\delta = 2$ and 4% when $\delta = 5$.

The cohort effects show a familiar pattern, even though the share of returns to education in income is higher for the 30-40 year old cohort, the contribution to inequality is the lowest for this group. When $\delta = 2$, it reduces welfare the same as a 19% decline in aggregate income would if income were distributed equally. This compares with a reduction of welfare of 22% and 25% for the full sample and the 50-60 year old cohort respectively. The contribution to inequality of the base levels is also less for this cohort than for the other two, 34% for the 30-40 compared to 38 and 43 for the full sample and the 50-60 cohort respectively. The effect of the interaction term in terms of increasing welfare is comparable between the two cohorts and the full sample. The effect naturally varies with the inequality aversion parameter but the same pattern emerges when examining the case when the inequality aversion is higher.

Table 4.9 confirms the results from the previous analysis. It appears that returns to education are of less importance to the inequality in earnings of the 50-60 cohort than they are for the other cohorts, as shown in column two. This difference is fairly small,
being only a difference of 2% between the cohorts when \( \delta = 2 \) but increasing to 4% when \( \delta = 5 \). The 30-40 cohort has the smallest contribution of education to measured inequality. Again compare the values in column 2 of table 4.9 with the 4% and 20% found by Taubman, and Layard and Zabalza. Inequality in the base level of income seems, from column one, to contribute between 60% and 70% of the inequality. The sample 30-40 has the highest value at 73% and 64% for \( \delta = 2 \) and \( \delta = 5 \) respectively, the full sample has values of 7% and 62% with the 50-60 cohort having values of 69% and 64%. The interaction effect reduces measured inequality by between 2% and 13%. The interaction effect again seems strongest for the 30-40 cohort and weakest for the 50-60 cohort. Slightly more variability is seen in the returns to the base level and the interaction effect than is present in the percentage contribution of the returns to education but the numbers are still very similar. Compared to the straightforward OLS estimates the percentage effect of education on earnings inequality is greater with the tobit model while the effect of the base level is smaller for this model than for the OLS model. This difference is uniform across cohorts and inequality aversion.

I again consider the marginal interactive decomposition for the tobit model of earnings. These results are given in table 4.10. From table 4.10, both the full sample and the cohort aged 30-40 reflect a similar pattern as in the OLS analysis. The major difference in these two samples is the magnitude of the effects. The marginal effects of education are greater with the tobit model while the effects of the base returns are less. A major qualitative difference appears when the full cohort is examined. In this sample both earnings sources have negative marginal contributions to inequality when \( \delta = 5 \), meaning that both of the earnings sources act as earnings equalizers when \( \delta = 2 \) however the effect of the base is to increase inequality as indicated by the positive sign on the marginal effect. Again it seems that the equalizing effect of education is much more important for the 30-40 cohort than for the other two cohorts. This is in contrast to the OLS analysis.
Chapter 4. The Contribution of Education to Earnings Inequality

where the base level of earnings was a disequalizing influence on total earnings.

The major conclusions from this subsection are that it appears that, when the mass point at zero earnings is not explicitly modelled, the share of returns to education in total earnings is under estimated and the actual inequality of returns to education will also be underestimated. The combination of the two effects results in an estimate of the total effect on inequality that is biased downwards towards a lesser effect of education on inequality, when the returns to education are estimated without explicit consideration of the change in the probability an individual will be employed.

4.7 Differing Education Effects

In the previous section of this paper, no attempt was made to determine how the contribution of returns to education to earnings inequality varied across the level of education. That is the subject of this section. Throughout this section I use the earnings decomposition technique from the tobit model where employment effects are explicitly considered.

The first step in this analysis is the decomposition of earnings into a part due to post-secondary education, a part due to secondary education, and the base level. To
Chapter 4. The Contribution of Education to Earnings Inequality

<table>
<thead>
<tr>
<th>Age</th>
<th>( \delta )</th>
<th>( I(\bar{y}^1) )</th>
<th>( I(y^*) )</th>
<th>( I(y^{PS}) )</th>
<th>( s^1 )</th>
<th>( s^* )</th>
<th>( s^{PS} )</th>
<th>( C )</th>
<th>( I(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5767</td>
<td>0.3422</td>
<td>0.0811</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.6492</td>
<td>0.4849</td>
<td>0.8495</td>
<td>0.5767</td>
<td>0.3422</td>
<td>0.0811</td>
<td>-0.0803</td>
<td>0.5289</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.9662</td>
<td>0.8956</td>
<td>0.9989</td>
<td>0.5767</td>
<td>0.3422</td>
<td>0.0811</td>
<td>-0.0452</td>
<td>0.8995</td>
</tr>
<tr>
<td>30-40</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5723</td>
<td>0.3363</td>
<td>0.0914</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.5972</td>
<td>0.3903</td>
<td>0.7905</td>
<td>0.5723</td>
<td>0.3363</td>
<td>0.0914</td>
<td>-0.0787</td>
<td>0.4666</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.9444</td>
<td>0.8082</td>
<td>0.9965</td>
<td>0.5723</td>
<td>0.3363</td>
<td>0.0914</td>
<td>-0.0565</td>
<td>0.8468</td>
</tr>
<tr>
<td>50-60</td>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6271</td>
<td>0.3067</td>
<td>0.0609</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.6829</td>
<td>0.6274</td>
<td>0.9201</td>
<td>0.6271</td>
<td>0.3067</td>
<td>0.0609</td>
<td>-0.0641</td>
<td>0.6175</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.9773</td>
<td>0.9716</td>
<td>0.9999</td>
<td>0.6271</td>
<td>0.3067</td>
<td>0.0609</td>
<td>-0.0220</td>
<td>0.9550</td>
</tr>
</tbody>
</table>

Table 4.11: Decomposition of \( I(y) \) for Earnings Decomposition

To perform this decomposition, I essentially use the technique used in section six except that I consider two counterfactual distributions of education. The first is the same as previously considered, I give everyone the lowest level of education possible, denote the resulting earnings vector as \( \bar{y}^1 \). The second counterfactual education distribution gives everyone with high school or less their actual education level. Everyone else is given the education level of some post-secondary, denote the corresponding earnings vector as \( y^{PS} \). The return to post secondary education is therefore \( y^{PS} = y - \bar{y}^{PS} \) and the return to secondary education is \( y^* = y^{PS} - \bar{y}^1 \). I then use these three vectors of contributions in my decomposition of the earnings inequality index. The results are given in table 4.11.

Column six shows that the share of earnings that is a return to post-secondary education is actually between 9\% for those aged 30-40 and 6\% and 8\% for those aged 50-60 and the full sample respectively. As can be seen from the table in column three, the returns to post-secondary education tend to be distributed much less equally than the returns to secondary education. Since many more people actually have secondary education than have post-secondary the number of people with zero returns to post-secondary is higher than the number of people with zero returns to secondary education. Thus the spike at
Chapter 4. The Contribution of Education to Earnings Inequality

<table>
<thead>
<tr>
<th>Age</th>
<th>$\delta$</th>
<th>$s^1 I(\tilde{y}^1)$</th>
<th>$s^* I(y^*)$</th>
<th>$s^{ps} I(y^{ps})$</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.3744</td>
<td>.1659</td>
<td>.0689</td>
<td>-.0803</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.5573</td>
<td>.3065</td>
<td>.0810</td>
<td>-.0452</td>
</tr>
<tr>
<td>30-40</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.3418</td>
<td>.1313</td>
<td>.0723</td>
<td>-.0787</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.5405</td>
<td>.2718</td>
<td>.0911</td>
<td>-.0565</td>
</tr>
<tr>
<td>50-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.4282</td>
<td>.1924</td>
<td>.0609</td>
<td>-.0641</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.6129</td>
<td>.2979</td>
<td>.0662</td>
<td>-.0220</td>
</tr>
</tbody>
</table>

Table 4.12: Amount of Measured Inequality Attributable to Different Sources

zero in the distribution of returns to post-secondary education is much larger than the spike at zero for the returns to secondary. The returns are most unequally distributed for the 50-60 cohort and are most equally distributed for the 30-40 cohort.

Column three of table 4.12 shows that inequality in the returns to post secondary education have the same effect on welfare as a loss of between 6% and 9% of aggregate income. There does not seem to be any cohort differences in this case, with all of the cohorts having approximately the same results. The largest impact of high school education is felt by the 50-60 cohort while the least effect is felt by the 30-40 cohort. The effect of post-secondary education compares to an effect equivalent to a loss of 13% to 31% for the effect of secondary education in column two and 34% to 61% for the returns to the base level in column one. The direct effects of secondary and post-secondary do not add up to the overall direct effect of education, given in the previous section, because with the additional income source the interaction effect has changed slightly.

Table 4.13 presents the percentage contribution of the various sources of earnings to measured inequality. Examining this table the most interesting thing is that, comparing column two to column three, the returns to secondary education contribute much more
to earnings inequality than do returns to post-secondary education in all cases. This is so even though the actual returns to education are distributed much more unequally for returns to post-secondary education. This result seems primarily because of the much higher share of earnings that returns to secondary education have relative to returns to post-secondary education. Because of this, inequality in the distribution of returns to secondary education should intuitively be more important to society than inequality in the returns to post-secondary education. Again it seems post-secondary education is most important for the 30-40 cohort, contributing 15% or 11% of measured inequality. It seems least important for the 50-60 cohort. No real pattern emerges when examining the percentage contribution of high school education.

In most cases the direct contribution of secondary education is at least twice that of the contribution of post-secondary returns. This suggests that for social policy purposes we, as a society, should be concerned with reducing inequality at the lower end of the education spectrum. Say for example implementing policies that encourage individuals to remain in high school instead of increasing the number of individuals who enter university. This would reduce the number of people who have not completed high school.
and, for those people, increase their income from secondary schooling. This would reduce
the measured inequality in the low education categories. This policy would probably also
increase the demand for post-secondary education, requiring more funding, but the pri-
mary target of social policy, if it is desirable at all, should be secondary education. The
next chapter provides more concrete evidence for this statement.

I conclude this section with an examination of the marginal effects on inequality of
the different levels of education. The marginal decompositions in this case are given in
table 4.14. Table 4.14 confirms what has already been seen. The marginal effect of both
types of education on measured inequality is to reduce inequality. The marginal effect of
secondary education is much greater in magnitude in all cases than the marginal effect
of post-secondary education. This provides further support for the statement that social
policy should target lower education levels.

Results by cohort show that the marginal effect of all types of education is most
important for the 30-40 cohort. This again indicates the importance of education for
lower experience workers.

The major conclusions from this section are that secondary education is a much more
important determinant of earnings inequality than is post-secondary. The largest effect


Table 4.14: Marginal Effects of Income Sources on Inequality

<table>
<thead>
<tr>
<th>Age</th>
<th>( \delta )</th>
<th>( s^1 I(\bar{y}^1) )</th>
<th>( s^s I(y^s) )</th>
<th>( s^{ps} I(y^{ps}) )</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.0035</td>
<td>-0.0314</td>
<td>-0.0005</td>
<td>0.5573</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.0005</td>
<td>-0.0178</td>
<td>-0.0006</td>
<td>0.9184</td>
<td></td>
</tr>
<tr>
<td>30-40</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.0155</td>
<td>-0.0325</td>
<td>-0.0008</td>
<td>0.4844</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0151</td>
<td>-0.0243</td>
<td>-0.0012</td>
<td>0.8572</td>
<td></td>
</tr>
<tr>
<td>50-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>-0.0304</td>
<td>-0.0160</td>
<td>0.0001</td>
<td>0.6638</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.0118</td>
<td>-0.0054</td>
<td>-0.0001</td>
<td>0.9723</td>
<td></td>
</tr>
</tbody>
</table>
of post-secondary education is felt by the younger 30-40 cohort. This likely reflects an increased importance of education, relative to other characteristics such as experience, for the the younger cohort. Marginal analysis confirms that education is more important for the younger cohorts.

4.8 Shapley Contributions

In this section I briefly consider the other decomposition of the AKS index that I introduced in chapter 2, the Shapley decomposition. The Shapley decomposition is given by

$$I(y) = \sum_{j=1}^{J} s^j \left[ 1 - \frac{\sigma_j}{\mu^j} \right]$$  \hspace{1cm} (4.121)

The terms in the Shapley decomposition provide a measure of the total contribution of source $j$ to inequality.

The values of this decomposition are presented, for the income decomposition that does account for employment effects, in table 4.15. For ease of comparison I also repeat in table 4.15 the values of the decomposition for the direct interactive decomposition from table 4.7. Column one gives the degree of inequality aversion, columns two and three show the respective contributions of the base level of education and returns to education respectively, column four gives the contribution of the interaction term, and column five presents the total measured inequality.

The last six rows of table 4.15 give the Shapley decomposition for the inequality indices. Since the Shapley decomposition decomposes the inequality index exactly, there is no interaction term for this part of the table.\(^{18}\) With this decomposition it appears that for all of the groups, and for both inequality aversion parameters, the returns to

\(^{18}\)Note that this is not the same as the technique of Taubman, and Layard and Zabalza who each attributed all of the interaction to a given source of earnings. This technique can be thought of as allocating part of the interaction to each earnings source.
<table>
<thead>
<tr>
<th>Type</th>
<th>Age</th>
<th>( \delta )</th>
<th>1 ( y^1 )</th>
<th>2 ( y_e )</th>
<th>Interaction</th>
<th>4</th>
<th>5 ( I(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>Full</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.3744</td>
<td>.2213</td>
<td>-.0668</td>
<td>.5289</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.5572</td>
<td>.3810</td>
<td>-.0388</td>
<td>.8995</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-40</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.3418</td>
<td>.1879</td>
<td>-.0631</td>
<td>.4666</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.5405</td>
<td>.3509</td>
<td>-.0445</td>
<td>.8468</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive</td>
<td>50-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.4282</td>
<td>.2483</td>
<td>-.0591</td>
<td>.6175</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.6129</td>
<td>.3632</td>
<td>-.0210</td>
<td>.9550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapley</td>
<td>Full</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.3412</td>
<td>.1877</td>
<td>-</td>
<td>.5289</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.5265</td>
<td>.3730</td>
<td>-</td>
<td>.8995</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-40</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.3056</td>
<td>.1610</td>
<td>-</td>
<td>.4666</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.4957</td>
<td>.3511</td>
<td>-</td>
<td>.8468</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.4359</td>
<td>.1817</td>
<td>-</td>
<td>.6175</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.6046</td>
<td>.3504</td>
<td>-</td>
<td>.9550</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.15: Comparison of Decomposition Measures
education contribute less to inequality than the returns to the base level. The returns to 
education contribute the least to the cohort aged 30-40 while the 50-60 cohort and the 
full sample have similar results.

It is quite obvious from the analysis in this section that the different decomposi-
tion methods in chapter 2 provide very different information. Of the three, the Shapley 
decomposition and the direct interactive both provide information about how much of 
observed earnings inequality is a result of returns to education and how much is a result 
of the base level of earnings. The difference is that, in the Shapley decomposition, I have 
succeeded in unambiguously separating and assigning any interaction effect to a partic-
ular income source. The marginal interactive decomposition shows how the observed 
inequality would change if the given income source is removed. The difference between 
the marginal and direct interactive decompositions is the base level of inequality which is 
used. In the direct decomposition the base inequality measurement, to which everything 
is compared, is zero while in the marginal decomposition the base level of inequality is a 
weighted sum of the inequality present in each source\(^{19}\).

In order to facilitate the comparison between the Shapley decomposition, the direct 
interactive, and the methods of Taubman and Layard and Zabalza, I have in table 4.16 the 
percentage contribution of the two income sources to overall inequality. This table shows 
that about forty percent of the measured inequality in earnings is attributed to education. 
Compare these numbers to the approximately forty percent from the direct interactive 
decomposition for all the cohorts and the six and four percent from the Taubman and 
the Layard and Zabalza techniques with this data respectively. The percentage analysis 
confirms that returns to education are least important, in the Shapley decomposition, 
for the 30-40 cohort. This is a seeming contradiction to the interactive decompositions.

\(^{19}\)Assuming only two sources of earnings. In general the base level is a weighted sum of the inequality present in the distributions \(y^{-j}\).
Chapter 4. The Contribution of Education to Earnings Inequality

<table>
<thead>
<tr>
<th>Age</th>
<th>δ</th>
<th>$y^1$</th>
<th>$y^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>59</td>
<td>41</td>
</tr>
<tr>
<td>30-40</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>59</td>
<td>41</td>
</tr>
<tr>
<td>50-60</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>71</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>63</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 4.16: Percent of Inequality by Shapley Decomposition

One way of thinking of this apparent contradiction is that the Shapley decomposition allocates less of the interaction term to the returns to education than it does to the returns to the base level of education. This is reasonable since, thinking back to the marginal interactive decomposition, returns to education had a relatively strong equalizing effect on earnings for the 30-40 cohort. The results for the full sample are similar to the results for the 30-40 cohort. The percentage contribution of education is highest for the cohort aged 50-60.

The similarity in results between the two methods I introduced is comforting especially given that the calculated interaction effect in the direct interactive decomposition is relatively small and thus the percentage contributions of the direct effect should be close to that in the non-interactive Shapley decomposition.

4.9 Conclusions.

In this paper I have investigated the decomposition of measured earnings inequality into an effect due to education and an effect due to other factors. The paper is an empirical application of the inequality theory developed in chapter 2. I determine the
Chapter 4. The Contribution of Education to Earnings Inequality

The contribution of returns to education to earnings inequality in Canada using the S-Gini inequality index. The major empirical findings are that returns to education directly contribute approximately one third to four tenths of measured earnings inequality. With the remainder a result of inequality in returns to personal characteristics other than education and an interaction term that serves to reduce measured earnings inequality. I also demonstrate that a significant portion of the returns to education arises from the increase in the probability of finding employment that is a result of higher education.

The next section investigated the differential effect that returns to different levels of education have on measured earnings inequality. I demonstrated that the returns to secondary education have a much larger effect on earnings inequality than do returns to post-secondary education. A possible interpretation of this result is that social policy should be aimed at lower education levels such as high school education rather than higher education. The comparison with the other decomposition methods shows that the Shapley decomposition, since it measures the total contribution of a source to inequality and not just the marginal contribution of that earnings source, gives results that are similar in spirit to the direct decomposition, which also measures the total contribution and not the marginal contribution.

The paper provides an empirical illustration of the human capital based theories of the income distribution, such as Loury (1981), by quantifying the amount of inequality that may be attributed to education. As well, the differential in the effect of post-secondary education verses secondary education has a specific policy implication. Viewed solely in terms of earnings inequality, the emphasis in public policy should be on policies, such as promotion of high school completion, that target lower education groups. It is important to note that this analysis is concerned only with the inequality of earnings and is silent on how best to affect the level of earnings in the society.
Chapter 5

Policy Simulations
5.1 Introduction

In this chapter I use the earnings generating model from the chapter three, and the inequality decomposition from chapter two to examine the effect on measured inequality, the marginal decomposition of measured inequality, and the EDE of some social policy changes affecting the quantity of education that individuals obtain. These can be contrasted to quality changes, where the quality of education obtained by the individuals who attend school is changed but without changing the actual number of people at different levels.

The analysis of chapter four in this thesis showed that there is a large difference in the contributions of education to earnings inequality depending on the level of education; returns to elementary and secondary education have a much greater impact on observed earnings inequality than returns to post-secondary education. It was suggested that this may indicate that a social policy targeting lower education levels would be more effective at combating earnings inequality than a general policy of encouraging education at all levels. This possibility is one of the policy options analyzed in this chapter.

I consider three specific types of policies. The first type, a low targeted policy, is designed to increase the educational attainment of the individuals in the lowest two education categories. A high targeted policy is one which encourages further education for individuals who have completed high school but do not have a university degree. A broadly based policy is one which encourages individuals of all education levels to increase their educational attainment.

The actual specifics of the policies are left open. I use as my starting point the effect of the policy on education levels. This causes a bit of a problem since without being specific about the administration of the policies I am unable to consider how administrative costs may affect the desirability of a particular option. To the extent that the administrative
costs are small compared to the cost of education, this may not be an important omission.

A more important omission is that I do not consider the effect of the opportunity cost of education. When an individual decides to purchase education for another year, one of the factors that would affect this decision is the possibility of quitting school and entering the labour force full time with the current education level. Since I do not consider opportunity cost in this analysis it must be recognized as a limitation of the applicability of the results.

The rest of the chapter is as follows. Section two briefly considers some frictions in the market for education which may justify the need for social policies. In section three I estimate the costs involved in increasing an individual education level by one category. Section four considers what I term constant population policies. These policies are ones for which the total number of people affected by the policy remain constant when moving from policy to policy. Section five considers constant cost policies, where the total cost of the policy is held constant. The last type of policy to be considered is one where the probability of any one person being affected, given that they are in a targeted group, are held constant. These are considered in section six.

5.2 Frictions in the Education Market

In this section I present some arguments that can be used to justify the need for social policies designed to encourage individuals to increase their education purchases.

In standard human capital models\(^1\) an individual regards education purchases as an investment in lifetime earning power. As an investor the individual purchases education up to the point where the marginal cost of purchasing another unit of education is equal to the marginal increase in lifetime earning power. If the social discount rate is the same

\(^1\)For example Heckman (1976) or Becker (1964).
as the private discount rate, if the individual has access to perfect capital and factor markets, and if the individual has all of the available information, then the optimal amount of education for the individual to acquire from societies point of view is the same as the privately optimal purchase. In this case any social policy which changes the amount of education purchased will result in a decrease in welfare. These three conditions are seldom met in reality.

Consider first the assumption of perfect capital markets for students. Perhaps the single greatest reason for capital market imperfections facing students is the lack of collateral. Unlike most investments the purchase of education involves nothing physical. In addition, individuals who are purchasing education frequently have little or no physical assets. This makes the provision of collateral on a student loan very difficult. Without collateral to help prevent default, the incentives to lend money for the purchase of education are greatly reduced. In response to the imperfect capital markets a variety of schemes for the provision of student loans have been proposed. It is schemes such as government sponsored student loans that I have in mind to reduce this problem.

The second major imperfection in the market for education is an information problem on behalf of the purchaser. Individuals who are making education purchase decisions are usually young. Frequently they do not have information about the consequences of their purchase decision or about how their circumstances may change in the future. This lack of information may lead to incorrect or myopic decisions regarding the amount of education to purchase. For example a job at minimum wage may look attractive to someone who is sixteen and living at home but not nearly so when trying to support a family. To combat this type of problem a policy could be undertaken to inform those making the decisions just what the likely consequences of their decisions will be in later life.

The last way in which an intervention in the market for education may be beneficial is in the presence of externalities to education purchases. If the education of individuals
interacts with the education of his or her co-workers, for example if it increases not only his or her productivity but also the productivity of co-workers, then it may be in societies interest to promote education beyond what is optimal when considering only private benefits to education.

Any of the reasons outlined above, plus many others, would provide adequate reason for a social policy intervention in the market for education. In reality all of them likely are present to varying degrees so this type of intervention is worth examining more closely.

5.3 Costs of Education

In order to properly discuss the effect of a given set of policies on social welfare, I will be considering the social net present value of a given policy option. This involves estimating the benefits as well as the costs of the policy discounted to the same base year. A policy which has discounted benefits greater than discounted costs is said to have a positive net present value and is socially desirable. The costs that I consider are only the direct costs of education. In other words I do not consider any dead weight loss due to increased taxation to cover the increased education expenditures. Living costs of the student have to be met whether or not a policy affects the education purchases and so are not included as a cost of education. A potentially important cost that I ignore is the opportunity cost of education. If an individual is pursuing education then he or she cannot be earning as much of an income as he or she could be if time spent in acquiring the education was instead spent full time in the labour force. The fact that opportunity costs of education are ignored is a limitation to the analysis.

I also adopt the simplifying assumption that the increased cost is financed by a proportional income tax. This has the effect of multiplying everyone's pre-tax earnings by a constant to arrive at post-tax earnings. Since the S-Gini AKS index is a relative index,
Chapter 5. Policy Simulations

<table>
<thead>
<tr>
<th>Level</th>
<th>Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary and Secondary</td>
<td>21,946,596</td>
</tr>
<tr>
<td>Post-Secondary Non-University</td>
<td>2,784,226</td>
</tr>
<tr>
<td>University</td>
<td>7,034,994</td>
</tr>
<tr>
<td>Vocational Training</td>
<td>2,814,113</td>
</tr>
</tbody>
</table>

Table 5.17: Government Expenditures on Education

this does not change the amount of measured earnings inequality from pre to post-tax earnings. In reality the Canadian tax and transfer system is commonly thought to be somewhat progressive, which means that after-tax earnings inequality will be somewhat overstated by the analysis presented here, compared to what would be the actual situation.

The first step in constructing the costs of education is to determine the expenditure on education. Table 5.17 shows the government expenditures on education for four categories. Government expenditures on elementary and secondary education are combined. I therefore use the technique of Constantos and West (1991) to separate the costs into an amount for elementary students and an amount for secondary students. An initial expenditure per student is obtained of 4,437 dollars a year. Secondary education is assumed to be 1.3 times as expensive as elementary education. Using this and the enrolment weights of .3653 for elementary enrolment and .6347 for secondary enrolment, I obtain an expenditure per student for elementary education of 4161 dollars and for secondary education of 5198 dollars.

I divide the expenditure on community colleges by the total non-university post-secondary enrolment to get a figure for the non-university, post-secondary expenditure per student of 10,631 dollars. To calculate an expenditure per student for universities, I again follow Constantos and West and assume that each part time university student

---

3These are calculated using the data in table 7 of Statistics Canada Education in Canada (1985-1986)
is equivalent to one half of a full time student. The subsequent calculation gives an expenditure per full time student of 14,207 dollars.

Table 5.18 gives the subsequent calculation of the total per student cost. The percentage of total cost born by the government is from Constantos and West. The only concern may be the use of the University figure for the other post-secondary. This is used to produce a conservative estimate of the cost of the education so that the use of the cost in the welfare analysis will in turn be conservative.

The costs of education are going to be compared to the welfare arising from lifetime income. Thus I compound the per year cost for the additional time spent in school. The additional time spent is determined by the following pattern. Category one takes 9 years, category two takes 12 years, category three takes 13 years, category four takes 14 years and category five requires a total of 17 years of education. I do the compounding at an assumed social discount rate of 3% and 6%. The resulting total cost of an additional step in education are shown in table 5.19

These estimated costs are in line with other estimates such as those in Constantos
and West (1991) and Vaillancourt and Henriques (1986). The assumed real interest rates of 3% and 6% are similar to the social rates of discount assumed in Layard and Zabalza (1976). Given the estimates of the cost of education I can continue with the policy simulations. Again it is worth reemphasizing that I have considered only the direct costs of education and have not considered the opportunity costs of education.

5.4 Constant Population Policies

In this section I consider policies that hold the number of individuals affected by the policies constant. I analyze three types of policies; a policy which increases the category of education by one with a probability of ten percent for the individuals who initially are in categories one to four, a policy which increases the category of education with a probability of 14 percent for those initially in categories one or two, and a policy which increases by one with a 33 percent probability for those in categories three and four. The probabilities are chosen so that the number of individuals affected by each policy remains approximately the same. Since the numbers of people in the various categories is different, this obviously indicates that different probabilities are needed.

Education is an investment which increases earning power over a persons entire working life so it is inappropriate to consider only one years earnings in analyzing policies pertaining to education. I therefore simulate, using the earnings model in the previous chapter, discounted lifetime earnings and do the welfare and inequality analysis using these discounted lifetime earnings.

Simulated lifetime income is calculated for a group in the population. I use the personal characteristics, except for age, of the 30-40 sample in the simulations. I use the personal characteristics of individuals in this cohort because any policy that a government initiates is likely to impact the younger members of society most heavily. This is especially
important for education policy since any impact of education policies will depend on the current distribution of education within the targeted groups. The 30-40 cohort is used since they are still relatively young, and thus approximate in personal characteristics the current group of individuals attending school, but are old enough so that major lifetime choices such as education have already been completed. For the simulations I have the first time period for everyone be when they are twenty-five.

In calculating the simulated incomes the first step is to use the personal characteristics of the age 30-40 cohort to calculate expected potential earnings from equation 4.116 for each individual in the cohort. I do this for each of thirty-five years starting at age twenty-five. I thus have, for each individual a vector of 35 yearly expected earnings. Then a random draw is taken from the normal distribution with mean 0 and standard deviation equal to the estimated standard deviation of the error in equation 4.116, for each individual, and added to the yearly potential earnings. This represents the individuals person specific characteristics, luck and other unmodelled determinants of earnings. Using the same error term for each year assumes that the whole estimated error term in equation 4.116 is fixed across time. It is not possible to use the actual estimated error term from equation 4.116 in the simulations since it is not observed for individuals with zero earnings.

At this point I have, for each individual, thirty-five years of potential earnings, representing an entire working lifetime. Then every negative value of potential earnings is set to zero to reflect the fact that the person is considered unemployed in that year. Thus every individual has simulated yearly earnings for each of thirty-five years which are either zero or positive. Finally the income in each year is decomposed in the same way as in chapter three, into a base level and a part due to education. The resulting vectors of earnings are discounted back to age twenty-five at a rate of 3% or 6% per year to give the discounted value of simulated earnings at age twenty-five.
Chapter 5. Policy Simulations

Table 5.20: Effect of Policies on Lifetime Inequality and Welfare

<table>
<thead>
<tr>
<th>Policy</th>
<th>$\delta$</th>
<th>$y^1$</th>
<th>$y^e$</th>
<th>$C$</th>
<th>$I(y^1)$</th>
<th>$I(y^e)$</th>
<th>$I(y)$</th>
<th>$\xi$</th>
<th>$\Delta \xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>423,436</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.021</td>
<td>-0.064</td>
<td>.512</td>
<td>.615</td>
<td>.433</td>
<td>.470</td>
<td>224,459</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.019</td>
<td>-.047</td>
<td>.871</td>
<td>.950</td>
<td>.809</td>
<td>.843</td>
<td>66,352</td>
<td></td>
</tr>
<tr>
<td>Broad</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>430,145</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.022</td>
<td>-0.067</td>
<td>.511</td>
<td>.615</td>
<td>.465</td>
<td>229,986</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.021</td>
<td>-.050</td>
<td>.868</td>
<td>.950</td>
<td>.801</td>
<td>.838</td>
<td>69,518</td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>428,873</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.025</td>
<td>-.068</td>
<td>.507</td>
<td>.615</td>
<td>.420</td>
<td>.64</td>
<td>229,780</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.023</td>
<td>-.050</td>
<td>.865</td>
<td>.8950</td>
<td>.796</td>
<td>.837</td>
<td>69,729</td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>432,814</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.017</td>
<td>-.068</td>
<td>.517</td>
<td>.615</td>
<td>.435</td>
<td>.466</td>
<td>231,193</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.015</td>
<td>-.050</td>
<td>.873</td>
<td>.950</td>
<td>.810</td>
<td>.839</td>
<td>69,622</td>
<td></td>
</tr>
</tbody>
</table>

I then assume that a S-Gini SEF defined over discounted lifetime income represents societies preferences and continue with the decomposition analysis from chapter three.

Table 5.20 shows the results of the lifetime analysis, using the marginal interactive decomposition of the AKS index, of the broad policy, the low targeted policy, and the high targeted policy.

Examining column seven of the table it is apparent that all of the policies reduce measured inequality of total earnings but that this effect is quite small, being only about a reduction of .01. The marginal effect of education with all policies is to reduce measured inequality as shown in column three. This decrease varies from -.03 to -.07. The marginal effect of the base level of income is to increase measured inequality by .02 to .03.

In order to perform a welfare evaluation for these policies I compare the change in the EDE income given in column six of table 5.20 with the average cost per person of the policies. The average cost per person of a policy is calculated by first finding the number of people affected in each education category, then multiplying by the cost per person of the increased education from table 5.19. This gives the total cost of the policy. For
Table 5.21: Calculation of Cost of low Targeted Policy

<table>
<thead>
<tr>
<th>Move</th>
<th>Prob.</th>
<th>Pop.</th>
<th>Cost per Move</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>.1384</td>
<td>287</td>
<td>17220</td>
<td>683,992</td>
</tr>
<tr>
<td>2-3</td>
<td>.1384</td>
<td>1787</td>
<td>10950</td>
<td>2,708,163</td>
</tr>
</tbody>
</table>

The cost of the policies targeted towards low education individuals is calculated in Table 5.21. The total cost of the low targeted policy is calculated by summing the two values in column 5 and is therefore 2,712,155 dollars. The high targeted policy costs 11,170,374 dollars, while the broad policy costs 5,885,960 dollars. Average cost per person is then the total cost divided by the total number of people, or 709 dollars, 2,923 dollars, and 1,540 dollars for the low, high and broadly based policies respectively.

The average cost per person can be compared to the increase in the EDE discounted income to see if the given policy is welfare increasing. If the increase in the EDE income is larger than the average cost per person, then the social net present value of the policy is positive and the policy is welfare enhancing. Since all of the policies have an increase in the EDE that is greater than the cost, they all have a positive net present value. The calculated net present value of the policies when $\delta = 2$ is 4818 for the broadly based policy, 1958 for the low targeted policy, and 2915 for the high targeted policy. Thus the result is that the policies can be ranked in terms of net present value with the broadly based policy most desirable, then the high targeted policy, and then the low targeted policy. When $\delta = 5$ the result is that the policies remain desirable but since the change in EDE is smaller, they are less desirable. The results when $\delta = 1$ are that all of the policies are more desirable than when $\delta = 2$, or $\delta = 5$. This is because none of the increase in individual earnings is wasted by inequality when there is no inequality aversion. In general, the more inequality aversion in the SEF, the less desirable any policies will be. Inequality aversion reduces the benefit of the policies but makes no difference for the
Chapter 5. Policy Simulations

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>$\delta$</td>
<td>$y^e$</td>
<td>$y^c$</td>
<td>$C$</td>
<td>$I(y)$</td>
<td>$I(y^e)$</td>
<td>$I(y^c)$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>None</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>282,488</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.022</td>
<td>-.064</td>
<td>.510</td>
<td>.616</td>
<td>.429</td>
<td>.469</td>
<td>150,094</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.020</td>
<td>-.047</td>
<td>.869</td>
<td>.950</td>
<td>.807</td>
<td>.842</td>
<td>44,477</td>
</tr>
<tr>
<td>Broad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>286,757</td>
<td>4,269</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.023</td>
<td>-.067</td>
<td>.508</td>
<td>.616</td>
<td>.422</td>
<td>.464</td>
<td>153,656</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.021</td>
<td>-.050</td>
<td>.866</td>
<td>.950</td>
<td>.799</td>
<td>.838</td>
<td>46,543</td>
</tr>
<tr>
<td>low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>285,962</td>
<td>3,474</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.026</td>
<td>-.067</td>
<td>.504</td>
<td>.616</td>
<td>.417</td>
<td>.463</td>
<td>153,527</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.028</td>
<td>-.050</td>
<td>.863</td>
<td>.950</td>
<td>.794</td>
<td>.837</td>
<td>46,683</td>
</tr>
<tr>
<td>high</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>288,496</td>
<td>6,008</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.018</td>
<td>-.067</td>
<td>.514</td>
<td>.616</td>
<td>.432</td>
<td>.465</td>
<td>154,469</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.017</td>
<td>-.050</td>
<td>.871</td>
<td>.950</td>
<td>.808</td>
<td>.838</td>
<td>46,636</td>
</tr>
</tbody>
</table>

Table 5.22: Effect of Policies on Lifetime Inequality and Welfare, Discounted at 6% costs.

I now repeat the above analysis but with a discount rate of 6%. Table 5.22 gives the results of the constant population analysis with a discount rate of 6%. The same general pattern is present here as with the discount rate of 3%. All three policies decrease measured inequality by a small amount. Again the marginal effect of education is to reduce inequality while the marginal effect of the base level is to increase inequality.

To do the welfare analysis I use average cost per person of 1,612 dollars for the broadly based policy, 919 dollars for the low targeted policy, and 3,081 dollars for the high targeted policy. Comparing this to the change in the EDE income gives the net present value of the policy change. For all policies but the high targeted one, the social net present value is greater than zero. Again the policies are most desirable when there is no inequality aversion present in the SEF. The net present value is negative for the high targeted policy, but in all other cases is positive, indicating that the benefits outweigh

---

4 These are calculated in an identical way to the previous values except the cost of education is compounded by 6% over the period required to change education levels.
Chapter 5. Policy Simulations

The constant population policies in this section are all, with one exception, worthwhile by the NPV criterion. A difficult arises when comparing them since the costs of the policies are not the same. For example the NPV of the high targeted policy, when \( \delta = 2 \) and the discount rate is 3% is higher than the NPV of the low targeted policy but the low targeted policy costs substantially less. Therefore the next section compares policies which all cost the same amount.

5.5 Constant Cost Policies

The policies that I considered in the previous section of this chapter all had the feature that the number of individuals who are affected by the policy change remained constant. Given that the policies affected the same number of people, this meant that the cost of the policies had to be different. In this section I hold the cost of the policies constant and change the number of people affected by the policy.

I consider only policies that cost five million dollars to implement and again concentrate on broadly based policies, policies targeting low education groups and policies targeting high education groups. The way that I calculate the lifetime income is identical to that in the previous section except for the probability that a given person in a group is affected by the policy. For the broadly based policy everyone in groups one to four has a probability of 8.5% of having their education level increased by one. The low targeted policy increases those with education categories one and two with a probability of 20.4%, and the high targeted policy increases those with category three and four with a probability of 14.5%.

The results of these simulations are shown in table 5.23. As can be seen in column five all of the policies again decrease the inequality in lifetime income. Very little change
occurs in the marginal components of earnings inequality in columns one and two. The pattern of education being equalizing and the base income being disequalizing is again present.

To examine the welfare effects of the policies I again must find the cost per person of the given policies. In this case it is easy as the total cost remains constant at five million dollars. With the 3,823 people I consider, this gives a cost per person of 1,308 dollars. Comparing this to the change in the equally distributed equivalent income shows that since all three policies result in an increase in EDE income which is greater than the average cost per person, all of the policies are desirable. The net present value of the policies can be ranked to show which policies give the best social return on the investment. The best is the low targeted policy, followed by the broadly based policy, and then the high targeted policy.

Table 5.24 gives the constant cost results for the case when the discount rate is 6% The same pattern occurs as when the discount rate is 3%; the return to education reduces inequality while the return to the base level increases inequality.
Table 5.24: Effect of Policies on Lifetime Inequality and Welfare, Discounted at 6%

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>$\delta$</td>
<td>$y^+$</td>
<td>$y^-$</td>
<td>$C$</td>
<td>$I(y)$</td>
<td>$I(y')$</td>
<td>$\xi$</td>
<td>$\Delta \xi$</td>
</tr>
<tr>
<td>None</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>423,436</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.021</td>
<td>-.064</td>
<td>.512</td>
<td>.615</td>
<td>.433</td>
<td>.470</td>
<td>224,459</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.019</td>
<td>-.047</td>
<td>.871</td>
<td>.950</td>
<td>.809</td>
<td>.843</td>
<td>66,352</td>
</tr>
<tr>
<td>Broad</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>286,091</td>
<td>3,603</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.023</td>
<td>-.066</td>
<td>.508</td>
<td>.616</td>
<td>.423</td>
<td>.465</td>
<td>153,170</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.021</td>
<td>-.049</td>
<td>.867</td>
<td>.950</td>
<td>.809</td>
<td>.838</td>
<td>46,252</td>
</tr>
<tr>
<td>low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>287,543</td>
<td>5,055</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.028</td>
<td>-.068</td>
<td>.501</td>
<td>.616</td>
<td>.411</td>
<td>.461</td>
<td>154,964</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.026</td>
<td>-.051</td>
<td>.860</td>
<td>.950</td>
<td>.789</td>
<td>.834</td>
<td>47,616</td>
</tr>
<tr>
<td>high</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>285,281</td>
<td>2,793</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.020</td>
<td>-.065</td>
<td>.512</td>
<td>.616</td>
<td>.431</td>
<td>.467</td>
<td>152,092</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.018</td>
<td>-.048</td>
<td>.871</td>
<td>.950</td>
<td>.808</td>
<td>.841</td>
<td>45,426</td>
</tr>
</tbody>
</table>

As far as the NPV calculation, the increased discount rate makes all of the policies less attractive. Similar to the constant population case, the high targeted policy when $\delta = 5$ becomes undesirable by the NPV criterion. The policies can again be ranked according to their NPV as the low targeted policy the best, followed by the broadly based policy, and then the high targeted policy is least desirable by the social NPV criterion.

5.6 Constant Probabilities

In this section I analyze a combination of the two previous types of policies. Each individual has a ten percent probability of receiving an increase in their education regardless of the number of people in his or her education category.

Table 5.25 shows the results of the lifetime analysis of the broad policy, the low targeted policy, and the high targeted policy given a discount rate of 3%. Examining column seven of the table it is apparent that the policies all serve to decrease measured earnings inequality compared to the status quo.
Chapter 5. Policy Simulations

Table 5.25: Effect of Policies on Lifetime Inequality and Welfare

<table>
<thead>
<tr>
<th>Policy</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>423,436</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>.021</td>
<td>-0.064</td>
<td>.512</td>
<td>.615</td>
<td>.433</td>
<td>.470</td>
<td>224,459</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.019</td>
<td>-0.047</td>
<td>.871</td>
<td>.950</td>
<td>.809</td>
<td>.843</td>
<td>66,352</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Broad</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>430,145</td>
<td>6,709</td>
</tr>
<tr>
<td>2</td>
<td>.022</td>
<td>-0.067</td>
<td>.511</td>
<td>.615</td>
<td>.615</td>
<td>.465</td>
<td>229,986</td>
<td>5,527</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.021</td>
<td>-0.050</td>
<td>.868</td>
<td>.950</td>
<td>.801</td>
<td>.838</td>
<td>69,518</td>
<td>3,166</td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>427,147</td>
<td>3,711</td>
</tr>
<tr>
<td>2</td>
<td>.024</td>
<td>-0.066</td>
<td>.509</td>
<td>.615</td>
<td>.424</td>
<td>.466</td>
<td>227,951</td>
<td>3,92</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.022</td>
<td>-0.049</td>
<td>.867</td>
<td>.950</td>
<td>.801</td>
<td>.840</td>
<td>68,531</td>
<td>2,170</td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>426,434</td>
<td>2,998</td>
</tr>
<tr>
<td>2</td>
<td>.019</td>
<td>-0.065</td>
<td>.515</td>
<td>.615</td>
<td>.434</td>
<td>.469</td>
<td>226,475</td>
<td>2,016</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.018</td>
<td>-0.048</td>
<td>.872</td>
<td>.950</td>
<td>.809</td>
<td>.842</td>
<td>67,312</td>
<td>960</td>
<td></td>
</tr>
</tbody>
</table>

The cost of the policies is calculated the same as in section three of this chapter. With a discount rate of 3% these are 641 per person for the low policy, and 899 per person for the high. Thus all of these policies are welfare increasing. It is interesting to note that the ranking of these policies, in terms of their NPV is the broad based policy, the low targeted policy, and the high targeted policy. Part of the reason for the dominance of the broad based policy in this case is that since everyone has a 10% chance of being affected, this policy affects substantially more people than do the other types of policies.

For a discount rate equal to 6% the results are given in table 5.26. The general effects when the discount rate is 6% are similar to those when the discount rate is 3%. The costs of the policies are calculated as 664 dollars per person for the low, and 948 dollars per person for the high policy. Using these numbers the NPV of all but the high targeted policy when δ = 5 is positive. The ranking by NPV is the same as when the discount rate is 3%.
5.7 Conclusion

This chapter uses the earnings generating model estimated in chapter three to construct simulated lifetime earnings distributions under several policy options. I then compare the policies in regards to their effect on earnings inequality and welfare. I show that in most cases, the policies increase the welfare of society more than they cost and are therefore socially worthwhile. Without exception, increasing inequality aversion makes the policies less attractive from society’s perspective because benefits are reduced by inequality aversion but costs of the policies are unaffected. For the constant population policies the ranking by NPV, from highest to lowest, is the broad based policy, the high targeted policy, and the low targeted policy. For the constant cost policies, the ranking is low targeted policies, broadly based policies, and high targeted policies. Finally for the constant probability model, the ranking is broad based policy, low targeted policy, and high targeted policy. My inclination is to say that the results for the constant cost

---

Table 5.26: Effect of Policies on Lifetime Inequality and Welfare, Discounted at 6%

<table>
<thead>
<tr>
<th>Policy</th>
<th>$d$</th>
<th>$y^k$</th>
<th>$y^p$</th>
<th>$C$</th>
<th>$I(y^k)$</th>
<th>$I(y^p)$</th>
<th>$\xi$</th>
<th>$\Delta \xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>423,436</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.021</td>
<td>-0.064</td>
<td>.512</td>
<td>.615</td>
<td>.433</td>
<td>.470</td>
<td>224,459</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.019</td>
<td>-.047</td>
<td>.871</td>
<td>.950</td>
<td>.809</td>
<td>.843</td>
<td>66,352</td>
</tr>
<tr>
<td>Broad</td>
<td>2</td>
<td>.023</td>
<td>-.067</td>
<td>.508</td>
<td>.616</td>
<td>.422</td>
<td>.464</td>
<td>153,656</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.021</td>
<td>-.050</td>
<td>.866</td>
<td>.950</td>
<td>.799</td>
<td>.838</td>
<td>46,543</td>
</tr>
<tr>
<td>low</td>
<td>2</td>
<td>.025</td>
<td>-.066</td>
<td>.506</td>
<td>.616</td>
<td>.421</td>
<td>.465</td>
<td>152,345</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.023</td>
<td>-.049</td>
<td>.865</td>
<td>.950</td>
<td>.799</td>
<td>.839</td>
<td>45,901</td>
</tr>
<tr>
<td>high</td>
<td>2</td>
<td>.021</td>
<td>-.065</td>
<td>.511</td>
<td>.616</td>
<td>.431</td>
<td>.468</td>
<td>151,394</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.019</td>
<td>-.048</td>
<td>.950</td>
<td>.808</td>
<td>.955</td>
<td>.841</td>
<td>45,101</td>
</tr>
</tbody>
</table>

---

Assuming that some sort of friction exists as in section two.
policies are the ones that are of most interest, since the results show where the social return on a dollar spent are highest.

This analysis has ignored the possibility of quality changes in the education that is purchased with no change in the number of individuals being educated. I ignore quality changes because firstly, quality of education is not something for which there is a well defined measure, and secondly, even if quality could be measured, it is very difficult to determine the effect of any change in quality on individual earnings.
Chapter 6

Conclusions
Chapter 6. Conclusions

Time and again level of education has been shown to be empirically related to labour earnings\(^1\). This thesis assumes that the relationship between the two is causal in that education determines earnings, and examines the effect that differences in education have on observed earnings inequality. The specific approach that I use is to measure earnings inequality with an S-Gini inequality index and determine how much of the observed value of the inequality index is a result of returns to education.

The first step in the analysis is a theoretical examination of how to decompose inequality indices by income source. The second chapter initially examines some previously known methods of decomposition and explains why they are unsatisfactory for the problem at hand. The chapter develops three new ways of doing the decomposition by income sources that are both theoretically consistent and intuitively appealing. The first two methods are termed interactive decompositions because of the distinguishing feature that the value of the inequality index is separated into direct effects for each income source plus an interaction term accounting for the fact that, in the absence of perfect correlation between the sources, inequality in one source of income has a tendency to counteract inequality in any other source. The third decomposition method introduced approaches the decomposition problem from a slightly different perspective. I ask that the values of the equally distributed equivalent income be divided into contributions for each income source, with no interaction term. I outline three intuitive properties that I want a decomposition to satisfy and use these to derive what I call the Shapley decomposition. This method is based on a decomposition of the equally distributed equivalent income function that is mathematically identical to the Shapley value of transferable utility cooperative games. The three methods are compared with each other as well as with the previously known decomposition methods. I conclude the second chapter with a brief discussion of possible uses of the decomposition methods in applied economics.

\(^1\)See Willis (1986) for a survey.
Chapter 6. Conclusions

In the fourth chapter of the thesis I proceed with one of the suggested applications. I first use the data on labour earnings from Statistics Canada's 1986 Family Expenditure Survey to estimate an earnings model. This model takes separate account of the probability of work and the income when working. I then use this estimated earnings model in a series of thought experiments which determine the effect of the distribution of educational attainments on the observed earnings distribution. I am able to divide labour earnings into two parts; a part due to returns to education, and a part that is the result of other personal characteristics. I use these two vectors of earnings sources in the interactive decomposition of the S-Gini index of relative inequality to determine the overall effect of education on earnings inequality.

An interesting feature of the earnings model used in this analysis, and one of the features that separates the analysis in this thesis from similar studies of the effect of education on inequality, is that the explicit modelling of the probability of an individual being employed allows the change in this probability associated with a change in education level to be one of the sources of returns to education. All else equal, a change in the amount of education will have two effects. The first is the increase in the individuals wage rate, and thus his or her earnings when working. The second effect is the increase in the probability that the individual will be able to find a job. Previous examinations of the sources of earnings inequality have considered the effect of the first of these factors but have ignored the second. One of the results of this thesis is that the change in the probability of working is a significant source of earnings, and of earnings inequality. The final results, including this effect, indicate that, depending on the sample used, and the degree of inequality aversion desired, between one third and one half of measured earnings inequality is a result of differing returns to education.

The fifth chapter of the thesis uses the earnings generating model presented in chapter three to conduct some policy simulations. This chapter differs from the previous ones in
that, instead of examining things year by year, I examine the effect of education on the inequality of lifetime earnings. I examine three types of social policies, policies targeting low education groups, policies targeting high education groups, and policies that target all groups equally. The model predicts that all of the policies I consider are in that the increase in costs of education more than outweighed any increase in welfare from lifetime earnings. No policy had large effects on measured inequality or on the amount of measured inequality that can be attributed to differences in education.

The questions raised in the empirical portion of the thesis are important for the insight that they provide to the more general problem of how best to reduce inequality. There is a rising interest among economists in the study of income inequality in the developed countries. This arises mostly out of a belief that inequality of incomes has been increasing and continues to increase\(^2\). This thesis has shown that a large proportion of earnings inequality on a year by year basis arises from differences in education. Thus education policy may be an important weapon in the fight against inequality\(^3\). Future research in this area may be fruitful, especially the use of panel data to develop true personal models of lifetime earnings and to use these lifetime earnings models in policy simulation experiments similar to chapter four.

\(^2\)See for example Jenkins (1992), Karoly (1992), and Johnson and Webb (1993).

\(^3\)This statement may seem to be in contradiction to the simulation results in chapter four but I don’t think that it is. The reason for this belief is that, while education policy changed inequality by only a small amount in absolute terms, the change in relation to the overall amount of inequality was fairly large. As well there is reason to believe that the amount of inequality over a lifetime is less than in a year by year comparison. Thus even though the examined education policies had little effect on lifetime inequality, the year by year effect may be more substantial.
Bibliography


[40] Statistics Canada: *Education in Canada (1985-1986)*.


Appendix A

Data Appendix
Appendix A. Data Appendix

In this appendix I describe the data set used in all of the analysis in the subsequent sections of the paper. I first describe the data set in general and then the actual variables that are used.

The raw data comes from the 1986 Survey of Family Expenditures compiled by Statistics Canada. The initial sample consists of 15334 observations on heads and spouses of households. These are designed to represent private households in Canada. The data set has some important exclusions that may be important. All residents of Yukon and Northwest Territories except for those living in Whitehorse or Yellowknife are excluded, all individuals living on Indian reservations are excluded, individuals in institutions such as old-age homes, penal institutions, and hospitals are excluded, and finally families of official representatives of foreign countries are excluded.

The data collected is given in table A.27. The data was screened as according to table A.28.

Table A.29 gives the sample means for the data that I use in the main analysis of the thesis where \( dl_i \) equals 1 the education level of person \( i \) is in category 1 and is 0 otherwise, \( age_i \) is the age of the head of the household in 1986, intended to proxy for experience, \( qu, on, sk, ab \) and \( bc \) are geographic dummies which equal 1 if the household is in the area, \( sex = 1 \) if the head of the household is a male, \( fr \) and \( en \) are language dummies which equal 1 if the first language is french or english respectively and finally \( m \) and \( s \) are 1 if the individual is married or single respectively.

Because I have only categorical data, I construct an education variable that is comparable to Taubman’s and Layard and Zabalza’s in the following way. Those with less than grade nine are assumed to have left school at 16, so they have \( ed = 9 \), those with some or completed high school have \( ed = 12 \), those with some post-secondary are given \( ed = 13 \), post secondary diploma, \( ed = 14 \) and university degree are given \( ed = 17 \) to account for those who took longer than average to get an undergraduate degree as well.
<table>
<thead>
<tr>
<th>Name</th>
<th>Possible Values</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wages</strong></td>
<td>[0,∞)</td>
<td>wages in dollars earned by the head of the household</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>20-65</td>
<td>age in years of the head of the household</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>0</td>
<td>female</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>male</td>
</tr>
<tr>
<td><strong>Weeksf</strong></td>
<td>[0,52]</td>
<td>weeks worked full time</td>
</tr>
<tr>
<td><strong>Weeksp</strong></td>
<td>[0,52]</td>
<td>weeks worked part time</td>
</tr>
<tr>
<td><strong>Self</strong></td>
<td>(0,∞)</td>
<td>Earnings from self employment</td>
</tr>
<tr>
<td><strong>Ed</strong></td>
<td>1</td>
<td>less than nine years</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>some or completed high school</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>some post-secondary</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>post-secondary diploma</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>university degree</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>not stated</td>
</tr>
<tr>
<td><strong>Geo</strong></td>
<td>1</td>
<td>Maritimes</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Quebec</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Ontario</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Manitoba and Saskatchewan</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Alberta</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>British Columbia</td>
</tr>
<tr>
<td><strong>Lang</strong></td>
<td>1</td>
<td>english</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>french</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>other</td>
</tr>
<tr>
<td><strong>Marr</strong></td>
<td>1</td>
<td>Married</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>single</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>other</td>
</tr>
</tbody>
</table>

Table A.27: Data Collected

<table>
<thead>
<tr>
<th>Screen</th>
<th>Number left</th>
<th>lost</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>15334</td>
<td></td>
</tr>
<tr>
<td>age ≥ 65</td>
<td>13003</td>
<td>2331</td>
</tr>
<tr>
<td>ed = 6</td>
<td>12929</td>
<td>74</td>
</tr>
<tr>
<td>age ∉ [30, 40)</td>
<td>3823</td>
<td>9106</td>
</tr>
<tr>
<td>age ∉ [50, 60)</td>
<td>2196</td>
<td>10733</td>
</tr>
</tbody>
</table>

Table A.28: Cuts to the Sample.
Appendix A. Data Appendix

<table>
<thead>
<tr>
<th>name</th>
<th>y</th>
<th>age</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>16761</td>
<td>40.511</td>
<td>.14289</td>
<td>.47843</td>
<td>.12374</td>
<td>.14158</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>name</th>
<th>d5</th>
<th>que</th>
<th>mt</th>
<th>sk</th>
<th>al</th>
<th>bc</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>.10767</td>
<td>.20788</td>
<td>.203015</td>
<td>.12705</td>
<td>.11174</td>
<td>.109905</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>name</th>
<th>fr</th>
<th>otl</th>
<th>sex</th>
<th>otm</th>
<th>s</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>.23218</td>
<td>.12412</td>
<td>.44582</td>
<td>.09651</td>
<td>.08852</td>
<td>-</td>
</tr>
</tbody>
</table>

Table A.29: Sample Means of the Data

as those with graduate degrees. I then construct an experience variable in the following
way \( ex = age - ed - 6 \).

Below is a listing of all of the variables that I use in the regressions and what they signify

- \( y \) yearly labour income,

- \( ed \), number of years of formal education. Described above.

- \( ex \) number of years of work experience. Described above.

- \( mt \) a dummy variable which is 1 if the individual lives in the maritime provinces

- \( que \) a dummy variable which is 1 if the individual lives in Quebec

- \( sk \) a dummy variable which is 1 if the individual lives in Manitoba or Saskatchewan

- \( al \) a dummy variable which is 1 if the individual lives in Alberta

- \( bc \) a dummy variable which is 1 if the individual lives in British Columbia

- \( sex \) a dummy variable which is 1 if the individual is male

- \( otl \) a dummy variable which is 1 if the individual’s first language is neither French nor English
Appendix A. Data Appendix

- $fr$ a dummy variable which is 1 if the individual’s first language is french
- $en$ a dummy variable that is 1 if the individual’s first language is english
- $otm$ a dummy variable for a person with marital status that is neither married nor single
- $s$ a dummy variable for a person who is single
- $age$ the age of the individual in years
- $d1$ a dummy variable which is 1 if the individual has less than nine years of schooling
- $d2$ a dummy variable which is 1 if the individual has some or completed high school
- $d3$ a dummy variable which is 1 if the individual has some post-secondary education
- $d4$ a dummy variable which is 1 if the individual has a post-secondary degree or diploma
- $d5$ a dummy variable which is 1 if the individual has a university degree
- $dlage$, $l \in [1, 2, 3, 4, 5]$ is an interaction term, the product of the education dummy $dl$ and the age variable.
- $mdl$, $l \in [1, 2, 3, 4, 5]$ an interaction term which is the product of the education dummy and the gender dummy
- $mdlage$, $l \in [1, 2, 3, 4, 5]$ the product between the gender dummy and the interaction between age and education
- $cfl$, $l \in [1, 2, 3, 4, 5]$ a dummy which is 1 if the counterfactual education category is category $l$
• $cflage, l \in [1, 2, 3, 4, 5]$ an interaction term between the counterfactual education level and age

• $mcfl, l \in [1, 2, 3, 4, 5]$ an interaction between gender and counterfactual education level

• $cflage, l \in [1, 2, 3, 4, 5]$ an interaction term between gender and $cflage$

• $\bar{y}^k$ is earnings when the base level of education is category $k$
Appendix B

Regression Results
Appendix B. Regression Results

In this appendix I present the results of all of the estimation in greater detail than in the text. All of the estimation is done with the complete sample as described in the data appendix. The variables used in the analysis are also described in the data appendix. Discussion of the results may be found in the body of the thesis where appropriate.

The first equation to be estimated is the replication of the Taubman analysis. The equation to be estimated is the following

\[ y = \beta_0 + \beta_1 ed + \beta_2 ex + \beta_3 ex^2 + \beta_4 mt + \beta_5 que + \beta_6 sk + \beta_7 al + \beta_8 bc + \beta_9 sex + \beta_{10} otl + \beta_{11} en + \beta_{12} otm + \beta_{13} s \]  

(B.122)

The equation is estimated using ordinary least squares. The results are in table B.30. The parameters and the associated standard errors are shown in the table followed by the \( R^2 \) of the regression, the standard error of the estimate, the log of the likelihood function and the number of observations used.

The second equation is the replication of the Layard and Zabalza estimation. Thus I estimate the equation

\[ f = \beta_0 + \beta_1 ed + \beta_2 ex \]  

(B.123)

where \( f \) is the ratio of the individual’s earnings to the mean earnings in the sample. The estimates for this equation are in table B.31. The results for the more general Layard and Zabalza equation which has more personal characteristics included are given in table B.32. These results are based on the following equation

\[ f = \beta_0 + \beta_1 ed + \beta_2 ed \times age + \beta_3 age + \beta_4 age^2 + \beta_5 mt + \beta_6 que + \beta_7 sk + \beta_8 al + \beta_9 bc + \beta_{10} sex + \beta_{11} otl + \beta_{12} en + \beta_{13} otm + \beta_{14} s \]  

(B.124)
### Appendix B. Regression Results

#### Table B.30: Estimates of the Taubman Equation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-27545.05</td>
<td>949.6</td>
</tr>
<tr>
<td>ed</td>
<td>2492.7</td>
<td>60.90</td>
</tr>
<tr>
<td>ex</td>
<td>1013.5</td>
<td>37.98</td>
</tr>
<tr>
<td>exs</td>
<td>-20.477</td>
<td>0.7472</td>
</tr>
<tr>
<td>mt</td>
<td>-4746.1</td>
<td>358.6</td>
</tr>
<tr>
<td>que</td>
<td>-2975.5</td>
<td>490.7</td>
</tr>
<tr>
<td>sk</td>
<td>-2600.5</td>
<td>404.0</td>
</tr>
<tr>
<td>al</td>
<td>-705.21</td>
<td>423.3</td>
</tr>
<tr>
<td>bc</td>
<td>-3117.4</td>
<td>423.6</td>
</tr>
<tr>
<td>sex</td>
<td>15030</td>
<td>236.6</td>
</tr>
<tr>
<td>otl</td>
<td>-1680.4</td>
<td>374.9</td>
</tr>
<tr>
<td>fr</td>
<td>35.472</td>
<td>452.4</td>
</tr>
<tr>
<td>otm</td>
<td>-105.79</td>
<td>404.5</td>
</tr>
<tr>
<td>s</td>
<td>-139.05</td>
<td>425.0</td>
</tr>
</tbody>
</table>

- $R^2$: .3806
- SEE: 13225
- LLF: -141051
- Observations: 12929

#### Table B.31: Estimates of the Layard and Zabalza Equation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-1.2662</td>
<td>0.0609</td>
</tr>
<tr>
<td>ed</td>
<td>0.1767</td>
<td>0.0042</td>
</tr>
<tr>
<td>ex</td>
<td>0.0025</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

- $R^2$: 0.1297
- SEE: .9340
- LLF: -17460.5
- Observations: 12929
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-2.96</td>
<td>.1900</td>
</tr>
<tr>
<td>$ed$</td>
<td>.1114</td>
<td>.0128</td>
</tr>
<tr>
<td>$ed \times age$</td>
<td>.0008</td>
<td>.0003</td>
</tr>
<tr>
<td>age</td>
<td>.10864</td>
<td>.0060</td>
</tr>
<tr>
<td>age$^2$</td>
<td>-.0014</td>
<td>0.00005</td>
</tr>
<tr>
<td>mt</td>
<td>-.2874</td>
<td>.0213</td>
</tr>
<tr>
<td>que</td>
<td>-.1785</td>
<td>.0292</td>
</tr>
<tr>
<td>sk</td>
<td>-.1525</td>
<td>.0240</td>
</tr>
<tr>
<td>al</td>
<td>-.0397</td>
<td>.0252</td>
</tr>
<tr>
<td>bc</td>
<td>-.1801</td>
<td>.0252</td>
</tr>
<tr>
<td>sex</td>
<td>.8926</td>
<td>.0141</td>
</tr>
<tr>
<td>otl</td>
<td>-.1106</td>
<td>.0223</td>
</tr>
<tr>
<td>fr</td>
<td>.0066</td>
<td>.0269</td>
</tr>
<tr>
<td>otm</td>
<td>-.0077</td>
<td>.0241</td>
</tr>
<tr>
<td>s</td>
<td>-.0078</td>
<td>.0025</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.3831</td>
<td></td>
</tr>
<tr>
<td>SEE</td>
<td>.7867</td>
<td></td>
</tr>
<tr>
<td>LLF</td>
<td>-141006</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>12929</td>
<td></td>
</tr>
</tbody>
</table>

Table B.32: Estimates of the Generalized LZ Equation.
For the main body of the analysis the following general specification is used which allows the return to education to be different for men and women.

\[ y = \beta_0 + \beta_1 d2 + \beta_2 d3 + \beta_3 d4 + \beta_5 d5 + \beta_6 d2age + \]
\[ \beta_7 d3age + \beta_8 d4age + \beta_9 d5age + \beta_{10} * md2 + \beta_{11} md3 + \beta_{12} md4 + \beta_{13} md5 \]
\[ + \beta_{14} md2age + \beta_{15} md3age + \beta_{16} md4age + \beta_{17} md5age + \beta_{18} age + \beta_{19} age^2 ex + \]
\[ \beta_{20} mt + \beta_{21} que + \beta_{22} sk + \beta_{23} al + \beta_{24} bc + \beta_{25} sex + \]
\[ \beta_{26} ot1 + \beta_{27} en + \beta_{28} otm + \beta_{28} \]  (B.125)

The estimation results for this specification are in table B.33. A restricted form of this specification, where the returns to education for men and women are restricted to be equal (ie coefficients \( \beta_1 \) through \( \beta_{17} \) are zero), is given in table B.34. The first two columns of both of these tables give the results when the equation is estimated by ordinary least squares. The last two columns give the results when the equation is given a tobit specification and estimated by maximum likelihood. Note that the values in the table for the tobit specification are the estimated values of the normalized coefficients, which are the estimates of the parameters \( \beta_k \) divided by the estimated standard error of the equation. Thus the reported coefficient on \( y \) is equal to the reciprocal of the standard error of the estimate.
Table B.33: Estimates of the Earnings Equation.
Table B.34: Estimates of the Restricted Earnings Equation.
Appendix B. Regression Results

Fig 1: Sample Censored Density

Fig 2: Density Estimated Ignoring Zeros

Fig 3: Density Estimated with Zeros
Appendix B. Regression Results

Lorenz Curves: Base and Education Income

1.2
0.8
0.4
0.2
0

% of Income

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
% of Population

Education ——
Counterfactual ——

Fig 5: Sample Indifference Curves