LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

( UNIVERSITY OF LONDON )

# Poverty, Intergenerational Mobility, and the Role of Imperfect Information: An Inquiry with Reference to the Panel Study of Income Dynamics

by

Ramses Hany Abul Naga

Thesis submitted for the degree of Doctor of Philosophy of the University of London

April 1994

الي جدتي العزيزة

و ذكري والدي الكّريم

*Memories brushing the same years.*

Paul Simon

## ABSTRACT

The standard approach to the study of poverty assumes the existence of an ideal variable that captures the extent of deprivation. In our first chapter, we postulate that poverty is involved with many dimensions . We use a latent variable framework to predict the extent of an individual's hardship as a function $\psi_i=ax_{1i}+bx_{2i}+...$, where the x's are indicators of i's income status, yi, and the latter variable is not observed.

In chapter 2, the problem of allocating benefits for poverty relief is considered in a situation of uncertainty about who the poor are. The decision to grant a benefit of fixed size is analyzed in the context of a social objective of minimizing poverty, subject to the social costs incurred by expenditure on poverty alleviation programmes. Chapters 1 and 2 contain empirical applications based on the Panel Study of Income Dynamics (PSID).

We then construct a theoretical model for the purpose of studying the relationship between poverty and credit market imperfections. These imperfections originate from information asymmetries between borrowers and lenders. We assess to what extent the above type of information asymmetry can be a cause of poverty. We identify cases where the presence of information asymmetries has a neutral effect, and other situations where it can cause reductions in the level of poverty.

We then review the issues pertaining to the topic of intergenerational earnings mobility. We define a criterion for the justice assessment of intergenerational mobility processes, based on the comparison of the earnings distributions of children originating from privileged groups and those from disadvantaged backgrounds. We apply our concept to intergenerational data from the PSID.

## ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor Anthony B. Atkinson for having taken the time and patience to supervise my work.

I also wish to thank Simon Bishop, Professor Frank Cowell, Dr Jean Dreze, Yiannis Doumenis, Dr Jean Duclos, Professor Karla Hoff, Dr Steve Howes, Dr Joanna Gomulka, Professor Steve Jenkins, Magda Mercader, Dr Danny Quah, Christian Schluter, and Dr. Jerzy Szroiter for helpful comments. I am especially grateful to Dr Georg Inderst, Marta Suranyi, Dr Dirk Van de gaer, and my father, for reading and re-reading various chapters of the thesis without hinting at the slightest feeling of fatigue, lassitude, or boredom.

This research was partly supported by the Swiss National Fund for Scientific Research. I also thank everyone at the Suntory-Toyota International Centre for Economics and Related Disciplines for their help, especially Brian Warren and Stephen Edwards.

Last but not least, I thank my extended family of friends and relatives in Alexandria, Cairo, Geneva, and London, for their moral support.

CONTENTS

---

TABLES AND FIGURES

*Tables*

*Figures*

CHAPTER 0

# Introduction and Summary

In this thesis we examine several topics in the economics of poverty and intergenerational mobility. Our first chapter deals with the problem of identifying the poor using multiple indicators of living standards. In our second chapter we examine the problem of state allocation of benefits when the needs of families are not directly observed by the decision maker. In the third chapter of the thesis we examine the extent to which credit market imperfections may influence the level of poverty in an economy with a dual sector structure. Our first two chapters deal with methodological questions, while the third follows a theoretical orientation, and provides a model of the determinants of income. Finally, in chapter 4 we propose a criterion for the justice assessment of intergenerational mobility processes.

Below we present in more detail the aims and contents of the various chapters, and how they all relate to one another. Three of our chapters contain empirical sections. They are based on the University of Michigan's Panel Study of Income Dynamics. In section 0.2 of this introductory chapter we describe our data. Then, in section 0.3, we discuss the data requirements for the study of intergenerational mobility, and we assess how close our data come to meeting these prerequisites. Though our data present several important shortcomings, we conclude that they offer some improvements over previous data sets, and thus that they may provide useful information regarding the topics we are dealing with in the thesis.

**0.1 Summary of the thesis**

We now summarize the four core chapters of the thesis and discuss how they relate to one another.

*Identifying the poor*

The standard approach to the study of poverty assumes the existence of an ideal variable that captures the extent of deprivation. The selected variable is often taken as income or consumption. The main problem with selecting one indicator at the expense of the other, is the well documented fact that alternative indicators may offer conflicting conclusions about who the poor are. In our first chapter, we discuss how several indicators of welfare can simultaneously be used in identifying the poor. We use a latent variable framework to predict the extent of an individual's hardship, as a function $\psi_i=ax_{1i}+bx_{2i}+...$, where the x's are observed indicators of i's income status, and the latter variable is not observed. The chapter contains an empirical application based on a sub-sample of the PSID, described in section 0.2 below.

*Allocation of benefits under uncertainty*

In chapter 2, we pursue the question of chapter 1 further. Given that it is not possible to exactly assess the needs of families, how should the state/decision-maker select families to be granted assistance? The decision to grant a benefit of fixed size is analyzed in the context of a social objective of minimizing poverty, subject to the social costs incurred by expenditure on poverty alleviation programmes. The chapter contains several empirical applications. Their purpose is to illustrate how changes in the key parameters of the decision problem, namely the poverty line, the size of the benefit, and the shadow cost of poverty relief policies, all influence the composition of the families which qualify for state support.

*Poverty and the economics of information*

In chapter 3 we construct a theoretical model for the purpose of studying the relationship between poverty and credit market imperfections. We consider an economy where individuals have to invest in their human capital in order to earn high wages. They may use the credit market in order to finance their training. The credit market is imperfect because the probability of default of borrowers is non-uniform, and unobservable to lenders. The recurring theme between this chapter and the previous two, is the relationship between poverty and imperfect information. The set-up however is somewhat different: in the first two chapters imperfect information occurs in a relation between agents and a government, whereas in chapter 3 imperfect knowledge arises in an exchange between private individuals (lenders and borrowers in the credit market). Also, we are no longer examining how to identify the poor, or how to allocate benefits in the light of imperfect information, but rather how imperfect information in the allocation of credit enters in the determination of income[(1)]. The aim of the chapter is to assess the extent to which the above type of information asymmetry can be a cause of poverty. We identify cases where the presence of credit market imperfections has a neutral effect, and other situations where it can cause reductions in the level of poverty.

*A justice criterion pertaining to intergenerational mobility processes*

One important conclusion that emerges from the model of chapter 3, is the importance of family background in the determination of income. Given that initial circumstances of individuals vary, how should we then judge the distribution of income at a point in time?

If 10% of the population today are poor, would it be irrelevant if these individuals all originated from low

---

(1) Imperfect information is an increasingly expanding area of microeconomics. The text of Hirschleifer and Riley (1992) offers a recent survey of the literature. Imperfect information also plays an important role in other fields. See for instance Laffont (1989) for public economics, Tirole (1988) for industrial organization, and Weiss (1991) for labour economics.

income families, or would it be a preferred state of affairs if they had purely random socio-economic backgrounds? In chapter 4 we attempt to answer the above question, where we place our focus on the normative assessment of intergenerational mobility processes. Looking at the joint distribution of incomes of parents and children, we ask the following question: what constitutes a just state of affairs? We define an intergenerational mobility process as being *un*just, if the income distribution of one group of children dominates the corresponding distribution for any other group of children, in terms of aggregate welfare. We apply our criterion to intergenerational data from the PSID, and find supporting evidence in favour of the injustice scenario.

**0.2 Data**

The empirical applications contained in the thesis draw evidence from the Panel Study of Income Dynamics, (PSID). The PSID is a U.S. nationally representative longitudinal survey of 5000 families, which has been running annually since 1968. The University of Michigan's Institute for Social Research provides annual documentation on the exact form of the questionnaire, definitions of variables, and the composition of the sample. The panel is particularly useful for the study of intergenerational mobility because the children of the original sample families interviewed in 1968 become automatically eligible respondents when they set up, or become members of, other family units. The survey thus allows one to gather information on the earnings of parents and children at similar stages of the life cycle.

The PSID is perhaps one of the most frequently analyzed data sets on incomes. This fact on its own makes it an attractive choice for empirical research. It often allows one to compare results with previous studies, and to get a feel for the plausibility of one's initial findings. The option of contrasting one's findings with those of related studies, is a substantial advantage for a graduate student with relatively little experience in empirical work.

The PSID being a specially conducted survey, generally offers wider information on family incomes and socio-economic backgrounds than data sets derived from administrative records (a typical source of which are tax declarations). Another factor which makes the PSID a valuable source of information resides in the fact that it is a prospective study. That is, it starts with the base sample, and updates the information on families at regular time intervals. This has meant that the survey centre has been able to answer the needs of researchers by collecting information on questions which have gained interest in recent years.

But not all is bright and clear. Use of the PSID calls for caution, especially with respect to the following three aspects: (i)its representativeness of the U.S. population, (ii) the non-response problem and, (iii) the attrition bias characteristic of long span longitudinal data sets. The sample selection and non-response problems are discussed below. The attrition question will be discussed in the next section where we assess the adequacy of the PSID for the study of intergenerational mobility.

In order to assess the representativeness of the PSID, one has to look into the history of the survey. In the years 1966-7, the Survey of Economic Opportunity (SEO) interviewed 30000 U.S. families laying particular emphasis on income, employment, and information on socio-economic status. The base sample of the PSID included 60% of its families newly selected from a national sample, while the remaining 40% were drawn from the SEO. The representativeness of the PSID is often questioned since the SEO fraction of the data over-sampled low income families living in urban areas(2) .

Because non-response to income surveys follows a non-random pattern, it is another source of sample selection bias. In a related U.S. study, Taubman and Wales (1974) find evidence that individuals with above average IQ and education are more likely to participate in surveys than

(2) The selected SEO families were chosen on the condition that their incomes did not exceed twice the US official poverty line.

those with relatively low IQ and education levels. The first round of the PSID had a response rate of 76%. However, this figure cannot be taken at face value, since 40% of the sample was extracted from the SEO, which at the time, had been running for two years. Thus, Atkinson, Bourguignon, and Morrisson (1992) pp. 49-50 write:

" In order to calculate the non-response, one has to take account of (i) non-response to the SEO in 1966 or 1967, (ii) refusal by SEO respondents to be given their name and address, and some losses in transmission of this information, (iii) non-response to the 1968 survey. Only the last of these is taken into account in the figure of 76% quoted above."

When examined over time, the non-response problem becomes that of sample attrition. We will return to this point in the next section of the chapter.

Our sample was extracted from the merged family-individual tape, of wave XX (1968-1987) of the PSID. The prime reason why we opted for a panel as opposed to a cross-section based data set, resided in our interest in examining questions pertaining to the topic of intergenerational mobility. Our first exploration of the PSID therefore consisted in constructing a data set for individuals participating in the labour force in 1986, whose fathers were present in the initial wave of the panel, and accordingly, who reported labour income for 1967. We called this sub-sample the *fathers and sons* data.

The initial sample comprised 20487 records and was reduced to 945 observations. Figure 0.1 details the steps involved with the selection procedure.

Our pairs of fathers and sons were selected in two main rounds. In the first stage we retained observations where the family was male headed, and where information was recorded on at least one dependent child.

Wave XX of the PSID: 20487 records

↓ Record kept if male, head or son, and if the family has at least one child.

3083 records

↓ Record kept if household head is the father and the child (children) is (are) son(s).

1458 (father,son) pairs

↓ Record kept if son's wage is recorded for 1986

968 (father,son) pairs

↓ Record kept if father's earnings for 1967 are reported.

945 (father,son) pairs

FIGURE 0.1 : Selection of the fathers and sons sample

From the selected families, the individual record was retained if the person was male. We were left with 3083 individual records. In the second round, we scanned these records in order to link fathers and sons. This stage was necessary since at several occasions, the household head was an older brother or a grandfather. We then deleted pairs if either the father was not working in 1967, or the son was out of the labour force in 1986, or if for either of them, information was missing for the earnings variable. As is shown in figure 0.1, it is mainly missing wages for sons as opposed to fathers which caused the final sample to shrink from 1458 to 945 pairs. Earnings data in the PSID are only recorded for family heads and wives. We therefore do not possess information on the earnings of sons who are still living in their parents home, regardless of their labour force participation status. We were left with a total of 945 pairs of fathers and sons. We had 609 distinct fathers, so that each father was given a sample weight equal to the number of sons he had.

There may be grounds for questioning the choice of focusing one's attention on fathers and sons exclusively. The debate on nature versus nurture cannot be separated from questions of inheritance of economic status. Following a *nurture* line, we may argue that genetic factors are irrelevant to studies of intergenerational mobility. Excluding a child who was raised by his mother's partner that does not happen to be the child's father would be a questionable decision. Within the *nature* stance, one could also raise objections to focusing our attention to fathers and sons exclusively. As the family structure has undergone many changes during the post-war era, it is no longer a rule that the mother stays home to raise the children, while the father is in full time work. In fact, reasoning along the lines of the *economics of the family* approach, we would conclude that the parent who has the higher earnings capacity would work in the market, while the other would specialize in housework. There may thus be good reasons to also include in our analysis mothers in full-time work, not just fathers only.

We used the fathers and sons sample throughout the thesis. The legitimacy of our choice cannot adequately be assessed prior to specifying the exact use, and aims of our empirical analyses. It is precisely for this reason that we devote the next section of the chapter to the assessment of the usefulness of the fathers and sons sample in the study of intergenerational mobility. Nonetheless we also use the sample in a different context, in chapters 1 and 2. There, we examine the identification and allocation of benefits to the poor, in the light of imperfect information. Since then we do not require information on the incomes of fathers, we carry out our applications only on the cross-section of sons. In chapters 1 and 2 we are illustrating empirical methods as opposed to testing hypotheses regarding population variables. In this sense, questions of population representativeness may appear of secondary importance when compared to the quality with which data are recorded, and the range of information available on the socio-economic condition of families.

## 0.3 Adequacy for the study of intergenerational mobility

In order to obtain an accurate picture of the degree of intergenerational mobility, it is generally perceived as important that: (1) evidence on the earnings of parents and children is gathered at similar stages of the life cycle, (2) that the sample is truly representative of the population we are studying, and (3) transitory variations and measurement errors in earnings are low. While the second and third points may follow from common sense, the first of these pre-requisites is less easy to justify.

Intergenerational samples can be derived from cross-section based data sources, as well as longitudinal, or panel, data sets. In the first of these sources, contemporaneous data on incomes are used, where parents are observed at a *mature* stage , and children at an *early* stage, of the life-cycle. Corresponding to the two stages of the life-cycle, denote parents incomes by $x_1$ and $x_2$, and children's incomes by $y_1$ and $y_2$. Cross-section samples, which we denote as $S_1$, contain observations of the type $(x_2,y_1)$. Longitudinal data sets contain observations such that the incomes of parents and children are ideally recorded twenty years or more apart. From longitudinal data, three other types of samples can be derived: $S_2$ containing observations of the type $(x_1,y_1)$, $S_3$ based on data of the type $(x_2,y_2)$, and finally $S_4$, a mixture of $S_2$ and $S_3$ data.

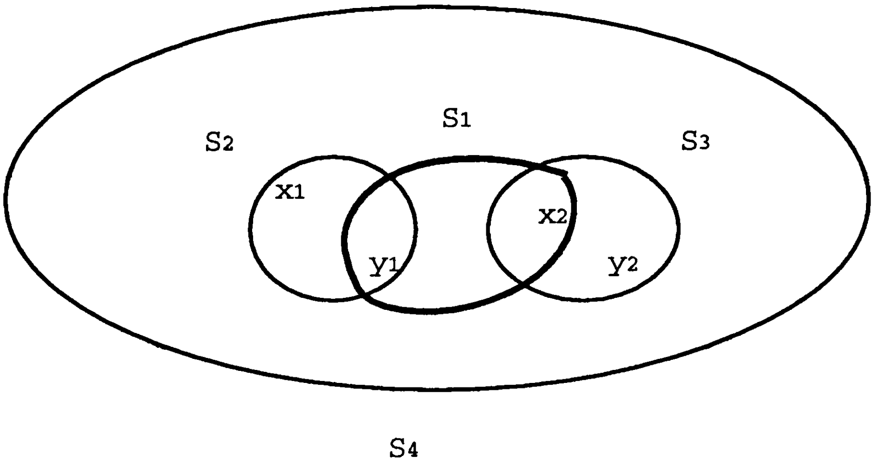


FIGURE 0.2: Types of intergenerational data

$S_2$ and $S_3$ have the distinct feature of containing observations recorded at an identical stage of the life-cycle, whereas in $S_4$, the two types $(x_1, y_1)$ and $(x_2, y_2)$ are present.

In practice, the choice between these four categories of samples is to a large extent arbitrary. As Jenkins (1987) explains:

"For each family sampled there is usually one observation for a parent and one for a child, rather than complete lifetime profiles. And because there are only these 'snapshots' observations rather than complete 'movies', observed status differences may not be indicative of 'true' intergenerational mobility."

The type of sample to be chosen to a large extent depends on what the researcher is interested in quantifying, and also on the assumptions made regarding the income-generating process underlying the data.

Focussing on the estimation of the regression coefficient between the incomes of fathers and sons, Jenkins has undertaken some simulations in order to compare the four types of samples. He concludes his study stating that:

"There is tentative evidence however that between-generation age differences are a greater source of bias than within-generation ones."

As with regards to $S_2$, $S_3$, and $S_4$:

"In any case, intuition is not a good guide to avoiding bias: same stage-of-the-life-cycle estimates are not necessarily better than contemporaneous ones, and purging earnings data of variation from within-generation age differences does not necessarily make estimates less biased."

We have therefore opted for an $S_4$ type of sample. By construction, $S_4$ contains $S_2$ and $S_3$ as sub-sets. taking 45 years of age as the boundary between the early and mature stages, the ratio of $S_2$ to $S_3$ observations is approximately 2:1 in our sample. From table 0.1, we see that the average age of sons is about ten years younger than that of fathers. The average age of a father was 39 in 1967, while the average age of a son was 29 in 1986. Our choice of an $S_4$ type of sample was essentially guided by the need of having a large enough data set. When necessary, observations of a specific nature were deleted in order to examine the likely impact of their omission on the conclusions previously reached.

| variable | mean | standard deviation |
|---|---|---|
| father's age in 1967 | 38.87 | 7.56 |
| father's wage in 1967 | 3.43 | 2.15 |
| Son's age in 1987 | 28.89 | 5.18 |
| Son's wage in 1987 | 10.8 | 7.76 |

Table 0.1: Ages and hourly earnings of fathers and sons

Another important pattern which comes out from table 0.1, is the seemingly higher degree of earnings inequality for the generation of children. For instance in taking the coefficient of variation (the standard deviation divided by the mean), k = 0.718 for children, while k=0.628 for fathers. The years 1967 and 1987 pertain to two distinct periods with regards to the evolution of earnings inequality in the United States. The period 1945-1973 was one of rapid earnings growth, and saw a doubling of real wages together with a moderate rise of earnings inequality. On the other hand, the period 1979-87 saw a rapid increase in inequality around an apparently stagnant mean (especially for men)[3]. The February 1992 issue of the Quarterly Journal of Economics contains several studies on the changes in the distribution of earnings in the U.S. during the eighties. The survey of Levy and Murnane (1992) devotes considerable attention to the vanishing of the middle classes, while the study of Hanratty and Blank (1992) focuses more on the evolution of poverty during the 1980s. On the whole, inequality increases were ascribed to *between* and *within group* factors. Between group factors appear related to the rise in the education and experience premiums. Within group factors (i.e. increasing pay differences between observationally equivalent workers) have been partly explained by phenomena such as increasing returns to skills, increasing industry specific wage differences, and also to plant specific factors within the same industry.

It does not automatically follow that the study of intergenerational mobility should focus on the earnings of parents and children as opposed to other variables such as family income or wealth. We will have more to say on this point in section 4.5 of chapter 4. At this stage though we should distinguish two main stances in the intergenerational mobility literature. One approach views parents as maximizing their utility over their own consumption and the children's welfare (typically Loury (1981)). The second approach made popular by the work of Becker (1981), assumes parents maximize their utility over their own consumption and their off-springs' income as opposed to his welfare.

(3) There is growing evidence that this observed trend may largely due to measurement errors involved with earnings surveys. See Heckman (1993) for a discussion.

The latter approach has received more attention than the former, and empirical work has mainly evolved around the theme of testing the inheritance of *economic success* as opposed to *economic welfare*. A consensus has then grown around the assumption that hourly earnings reflect best individuals economic success, so that most recent empirical studies today consist of examining how earnings of parents and children correlate (for e.g. Behrman and Taubman (1990)). Nonetheless if we wish to investigate the association between the welfare levels of parents and children, as opposed to the inheritance of economic success, the choice of income over earnings would leave little room for dispute. A family income-to-needs index would then be a more accurate indicator of family welfare than an earnings variable (in this context see Solon (1992)).

We now turn to the question of population representativeness. The PSID being a national probability sample, it is more likely to meet the requirements of condition (2) than in the case of previous studies. For instance the sample analyzed by Atkinson et al.(1983) was limited to low income parents living in the city of York. York is hardly representative of the U.K., nor are low income families a random sample in terms of socio-economic characteristics. also, the study of Behrman and Taubman (1985) was based on a sample of white male twins who served in the armed forces. While the PSID started as a representative sample of the U.S. (though see the discussion above), its reliability over the years has been hindered by the inevitable attrition problem, common to longitudinal data sets.

The attrition problem in the PSID was the subject of the study of Becketti et al. (1988). They reported a tendency for the panel to lose low income and high income individuals, so that the middle income classes became over-represented with the years.

"Individuals in SEO sample households are more likely to leave than those in SRC [non-SEO families] sample households, even after controlling for income. Low-income and high-income individuals are more likely to leave than those in the middle income categories. Sample [non-SEO] individuals collecting welfare in the last period are more likely to leave."

They concluded though that the attrition bias was quite moderate, though "the issue of the randomness of attrition should be considered carefully by each user of the data".

With regard to the third prerequisite stated above, it has become common practice to discard observations relating to individuals below the age of 25, because their current earnings were unrepresentative of their long term earnings profiles. There is also the factor of measurement error in the reporting of earnings and employment status, which tends to be unusually high amongst low skilled individuals and young workers(4). Since 98% of fathers , and 83.4 % of sons were above 25, our data set did not appear to be highly faulty in this context. Lillard and Willis (1978) have estimated in the PSID that the transitory variations were an important component of observed earnings. The auto-correlation coefficient of an individual's earnings over a six year period was of the order of 0.7. Thus there appears to be a substantial level of *intra*-generational mobility which is bound to put limits on the inferences that we may wish to make about the degree of mobility between generations.

Finally, before we embark on a study of this nature, it is wise to caution against making any strong conclusions on the basis of earnings data. In chapter 4 we will discuss the choice of the appropriate indicator of life-time income. Whether we select annual income, earnings, or hourly wages, we are severely limited by the inaccuracy with which earnings data are reported. The precision of income data is dependent on the accuracy with which earnings are reported. Likewise, hourly wage data will only be reliable if annual earnings and work hours are correctly reported.

---

(4) See the evidence in Bound et al. (1989).

The cross-validation study of Bound et al. (1989) consisted in comparing earnings reported by a sample of workers, with the data available form company records. It has shed important light on the nature of measurement error in the PSID. The usual assumptions that measurement errors have zero means and are uncorrelated with *true* values (i.e. company records), do not find any comfort when confronted with empirical evidence. Furthermore, the amount of measurement error appears considerable. The correlation between (log) reported earnings and true values is 0.81. This figure further drops to 0.56 for (log) hourly earnings. For (log) work hours, the correlation between reported values and company records is about 0.64. Also, workers with below average annual earnings tend to over-report their earnings, while workers with above average annual earnings tend to under-report their pay.

From the discussion above, it comes out that our data set is far from meeting the ideal conditions for the study of intergenerational mobility. Nonetheless, to the extent that parents and children are observed twenty years apart, that the attrition problem is inevitable in studies which span over such a long time interval, and that a large majority of fathers and sons were above 25 years of age at the time the data on their earnings were collected, and finally that wave XX of the PSID has not been the subject of a study of this nature, we found it worthwhile to carry out the data analysis, of which the main findings are contained in the next chapters.

CHAPTER 1

## Identifying the Poor: a Multiple Indicator Approach

### 1.1 Introduction

Two important aspects of any empirical study of poverty are what are known as the identification and aggregation issues. The identification rule dictates how we decide who is poor and often, but not always, it deals with the question of how poor the person is. The aggregation step enables us then to take the individual poverty data, and to summarize them into an economy-wide measure of poverty. For instance when we use the head-count to study the incidence of poverty, we label a person as poor if he falls under the poverty line. The head-count measure is the proportion of the population that is below the specified poverty line.

In this chapter we will be concerned with the identification problem in face of imperfect information. Assume that underlying our study of poverty lies a social welfare function $W(y_1,...y_n;\ y^*)$ where $y_i$ denotes the income status of family i, and $y^*$ is the poverty line. The identification rule then consists of separating families into two groups. The group for which $y<y^*$ is that of the poor. The families for which $y \geq y^*$ are members of the second group; the non-poor. The problem we are dealing with in this chapter is that of the identification of the poor when y is not observed, or is subject to measurement error. The income status of a family may be measured with error because it is systematically under-reported. If families may benefit from various welfare programmes when their incomes are low, they may have an incentive to understate their resources. Glewwe(1990) adopts this point of view in his work on the "efficient allocation of transfers to the poor". There are of course other reasons why y may be unobservable in practice. If the pertinent variable to welfare analysis is the long run economic status of the family ( e.g. its

permanent income ), then one must recognize that static indicators such as current income and consumption, may be noisy correlates of long run income status. In our work we will adopt this long run income, interpretation of y.

When y is unobserved, one may attempt to isolate its least noisy indicator, as Anand and Harris (1991) and Chaudhuri and Ravallion(1992) have done. However, one cannot ignore the fact that the most commonly used indicators of well-being, namely income and consumption, often offer different information about the composition of the poor: some families may cross the poverty line in the income space but not in the consumption space, and vice versa. It is not clear though why we should not simultaneously exploit information about various indicators of welfare. If, let us say, consumption is a good indicator of permanent income, is it reasonable to assume that it exhausts all the necessary information about the latter variable?

The aim of this chapter can be stated as follows: starting from a set X of p observed indicators on the family's socio-economic characteristics (for example: income, consumption, and employment situation of head), we want to construct a summary statistic of the family's long run income status y . The task of constructing such an index can be stated in the language of multivariate statistics, as one of reducing the dimensionality of the data from p to one dimension. As this is not a new problem in statistics, we have a vast literature and many results to draw from.

The index we select must achieve the reduction of the dimension of the data, without losing any available information contained in X, about y. That is, the index must be a *sufficient statistic* for y. One such statistic we propose to use is the regression function of y conditional upon X. Below, we will refer to it as the *multiple indicator index*.

To arrive at the multiple indicator index, we will proceed in the following steps. In section 1.2 we discuss why the two rival indicators of welfare, namely income and consumption, are not likely to rank families in terms of their well-being in an identical fashion. We also take a first look at our income and consumption data, and note the existence of conflicting conclusions offered by the two indicators. In section 1.3, we postulate the statistical relation between the indicators X and the unobserved income status of the family, y. We assume that the correlation between the various indicators arises out of their common dependence on y. This formulation leads us to adopt the model of *factor analysis* in describing the relation between the indicators and the unobservable. Within the context of a simplified example of the life-cycle model, we discuss the inferential problems involved with the factor analysis model. We then derive the multiple indicator index, which enables us to summarize the information about y.

In section 1.4 we illustrate the multiple indicator approach using a sample of 910 families, extracted from wave XX of the Panel Study of Income Dynamics. We then construct two classification matrices, where observations are ranked using the multiple indicator index, and each of the two commonly used indicators; consumption and income. These matrices offer greater scope for agreement concerning the ranking of families, than in the case where observations are ranked on the basis of income and consumption. We therefore summarize the results obtained from the suggested method, as being a "middle of the road" solution as an alternative to working with either of the two indicator separately. In section 1.5 we comment further on the use of the multiple indicator method, especially with respect to the limitations of its applicability. Section 1.6 concludes the chapter.

## 1.2 The Problem

As a first step to any study on the extent of poverty we have to specify the choice of our indicator variable of welfare. Let x denote such an indicator, and x* the specified poverty line. Should we define the standard as an income measure $x_1$ or should it be expenditure $x_2$? Conflicting conclusions can arise when one indicator is chosen at the expense of another. Consider first the case when $x_1 < x_1^*$ and $x_2 > x_2^*$. In Crime and Punishment, Dosteyevsky writes the following:

". . . possibly what weighed with her most was 'the poor man's pride' which makes poor people who are faced with the necessity of observing certain of our traditional customs strain every nerve and spend their last savings so that they should be 'as good as everybody else' and that no one 'could have a wrong word to say against them'".

Thus, a family may be able to reach a decent level of expenditure by temporarily borrowing or running down its assets. On the other hand, a situation where $x_1 > x_1^*$ and $x_2 < x_2^*$ can occur when there exist market imperfections due to information problems, discrimination, and other obstacles to trade. A well-off black family may afford a rent in a white suburb , but may be obliged to live on the other side of the railroad tracks due to the reasons mentioned above.

The need for examining multiple indicators of welfare has long been recognized in development economics. For instance, in their study of economic mobility and agricultural labour in Rural India, Dreze et al.(1992) write the following:

"One may however question whether current per-capita income in any particular year is a sensible criterion of 'poverty' in economies where current incomes are subject to large short-run variations and significant mechanisms exist for smoothing out these fluctuations. On the basis of alternative criteria of poverty such as per-capita expenditure or living standard, it is likely that less mobility would be observed. Households which may appear to be "crossing the poverty line " in particular years in the income space may, in fact, be chronically poor in terms of expenditure or living standards."

Also, Glewwe and Van der Gaag (1990) test on L.D.C. data the consequences of using various definitions of deprivation in identifying the poor. They conclude that the choice of a resource variable matters significantly , in the sense that different definitions identify different groups of the population as poor. The problem of conflicting results subject to the use of different indicators is not confined to the case of developing countries alone. Mcgregor and Boorah (1992) report substantial rerankings of individuals using income and expenditure data for the U.K. .

It is worthwhile asking the question as to when one can safely use one indicator instead of another; i.e. , under what circumstances is it that the problem of selecting a specific indicator amongst several, ceases to be a "problem". Let us once more restrict ourselves to the case where the choice is between two indicators: income ($x_1$), and consumption ($x_2$). We also abstract from the borrowing and other quantity constraints mentioned earlier. Then,

(i) if the correlation between $x_1$ and $x_2$ is equal to unity, one can use the information on either indicator to identify the poor.

(ii) one could weaken the condition (i), in requiring that the ordering of individuals by $x_1$ and $x_2$ be preserved. The choice between $x_1$ and $x_2$ would be unimportant provided the rank correlation between the two variables were unity.

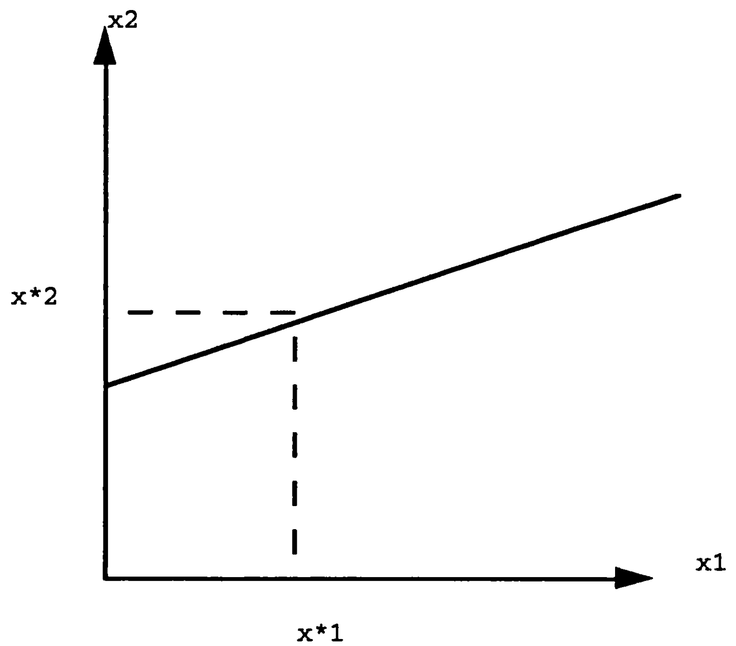


FIGURE 1.1: Income and consumption poverty lines.

One could then also convert the poverty line in terms of $x_1$ into a poverty line in terms of $x_2$.

The Keynesian consumption function is a typical example of condition (i) above. With a simple linear relation of the type $x_2 = a+bx_1$, the income poverty line $x^*_1$ is mapped into a consumption poverty line $x^*_2=a+bx^*_1$, leaving the identification of the poor invariant to the choice of indicator.

The linear consumption function is not likely to be of much relevance at the microeconomic level, unless we have a good reason to assume that savings do not respond to the interest rate. As mentioned above, we need not restrict ourselves to a linear relation between income and consumption, any one-to-one relation will serve our purposes. Deaton(1992) ch. 1, offers several reasons why one cannot expect consumption to closely follow income. One explanation relates to the fact that the marginal utility of consumption may vary along the stages of the life-cycle, especially with respect to the demographic structure of the household. More importantly, when we incorporate uncertainty in the course of the life-cycle, we note that the consumption patterns of households may substantially vary because of their different abilities to bear risk. As Deaton (p.19) very well summarizes the fact:

"In an uncertain world, the substitution of future consumption for current consumption inevitably increases exposure to risk, and those who are willing to contemplate the former must be willing to face the latter."

Life-cycle income would then be a key variable in the ability to bear risk. Looking at the relationship between current income and consumption only, would leave us with an incomplete, and possibly, inaccurate picture. A priori therefore, we cannot assume a one-to-one relation between consumption and income. We would thus expect a substantial reranking of families according to which of the two indicators we use.

*Data*

At this stage, it is worth confronting the evidence by taking a first look at the data we will be making use of in this chapter. We have data on income to needs ratio, food budget share, as well as employment status for 910 American families. These observations were extracted from the Panel Study of Income Dynamics, and pertain to 1986. The families are male-headed and constitute the cross-section of the second generation from the Fathers and Sons sample we shall be working with throughout the thesis. There are 945 families in the data set, but in the applications below we lose 35 observations because of missing values for the consumption variable. The point to note here is that the families retained all have a head who is participating in the labour force. Thus, individuals in full time education, elderly families, and other households who are not in the labour force are excluded from the analysis. The consequence of this selection rule is that families for which income is low and consumption is (relatively) high, may be under-represented.

The income variable is defined as the annual taxable income of the family, standardized by the Orshansky scale[(1)]. The food budget share is constructed as the ratio of annual food expenditure to the taxable income of the head and wife. Annual expenditure on food excludes food purchased with food stamps, and is limited to purchases for domestic use only. Families are asked the question "How much do you (or anyone else in your family) spend on food that you use at home in an average week?" The figure is then multiplied by 52 to arrive at the annual food consumption value. One should expect a substantial amount of inaccuracy to be involved with such data: respondents are likely to give more or less reliable information depending on how often they do the shopping, and also on the time horizon they use in estimating the average weekly expenditure on food.

---

(1) The Orshansky scale is based on the annual food requirements of families. Per capita needs in a family comprising four individuals is the reference point, say 100. Then, for a single person, p.c. needs rise to 120. The figure drops to 110 for two persons, 105 for three, 95 for five, and 90 for six or more individuals.

Ideally we would want to use an expenditure to needs ratio for the consumption indicator. However the PSID does not possess as refined data on consumption as the ones available on income. Noting that food consumption is typically an inferior good (at least in developed countries), a negative correlation between the income to needs ratio and food budget share is to be expected. In our data, it is of the order of -0.3. In the matrix below, we classify our observations line-wise in order of increasing income, and column-wise in decreasing order of food budget share. We adopt a quartile classification, so that income and consumption classes are of equal size, and contain exactly one quarter of the families.

$$T(x_1, x_2) = \begin{bmatrix} 138 & 51 & 23 & 15 \\ 65 & 88 & 51 & 24 \\ 20 & 66 & 86 & 55 \\ 5 & 22 & 68 & 113 \end{bmatrix}$$

By examining the off-diagonal elements of the matrix, one notes the extent of divergence in the classification of observations by the two indicators. A $\chi^2$ test of independence between $x_1$ and $x_2$ is quite expectedly rejected at 99%. While the indicators do not rank the families identically, there is a correlation between classifications obtained by the two variables (this result is not surprising since the correlation between $x_1$ and $x_2$ is -0.3). The extent of disagreement between the two indicators in the ranking of families can be quantified by counting the ratio of off-diagonal elements to the total number of observations. There are 465 out of 910 families ranked differently by $x_1$ and $x_2$, thus over 50% of the total. By setting the poverty line at a particular level for each indicator, we may of course observe that agreement between the two indicators may be higher or lower for this subset of individuals, than for the entire population. On the other hand, there is no general agreement on how, and at what level to set the poverty line. It would therefore appear more reasonable to discuss the adequacy of specific indicators in the applied analysis of living standards independently of where we may wish to set the poverty line. We will thus examine the performance of the indicators on the entire range of the population, rather than limit ourselves to a specific subset.

Nonetheless, as a simple exercise, assume that the poverty line is set on a relative basis, so that the bottom 25%, ranked according to a specific indicator, are defined as the poor. Then, from our classification matrix T we see that 138 families are ranked poor by both indicators; whereas 227 (i.e. 902/4) by construction, are identified as poor using each of the income and consumption indicators. In 40% of cases, the most popular indicators, income and consumption, identify different families as being in poverty. This result is largely in conformity with many economic theories of consumption. Since income is correlated with consumption, one indeed expects that the classification using the two indicators does not exhibit a pattern of independence. On the other hand , since the relationship between current income and consumption is not one-to-one, the ranking of individuals according to the two indicators is not likely to be uniform. We thus conclude this section by stating that indeed the choice between indicators in identifying the poor is a "problem".

### 1.3 A Multiple Indicator Approach

The evidence that alternative indicators offer different information concerning the identification of the poor, can be interpreted in three different ways. The first argument would be to say that there exists a proper indicator of welfare, say income, and to postulate that consumption is related to it through an economic law of the type:

$$x_2 = a+bx_1 + \varepsilon$$

where $\varepsilon$ is a random term. The researcher is then entitled to prefer $x_1$ to $x_2$ . This is probably the status-quo amongst researchers today, divided between those who favour the choice of consumption, and those who opt for income. Chaudhuri and Ravallion(1992) write:

"Current consumption expenditure and current income have been the most popular welfare indicators in applied welfare analysis. Of the two, current consumption expenditure has probably been more widely used for research and policy purposes, at least in developing countries."

An alternative approach is to argue that both income and consumption are correlates of an unobserved variable, which is pertinent to welfare analysis. For example, if our main concern is with chronic, rather than transient poverty, one would expect current income and consumption to be noisy measures of long term economic status. This time, the preference for $x_1$ over $x_2$ could be motivated on the basis that $x_1$ is a noisy indicator of a variable y, defined as long term economic status, and as before, $x_2$ is a noisy indicator of $x_1$. The problem arises because y typically is not observable. For example, if y is permanent income, there is no direct way of measuring it. Thus, permanent income is to be treated as a latent variable. Let us write a simple system of equations relating $x_1$ ,$x_2$, and y.

$$x_2 = a + bx_1 + w$$

$$x_1 = y + e$$

We see that $x_2$ will also be a noisy indicator of y. Substituting $x_1$ into the equation for $x_2$, we get:

$$x_2 = a + by + v$$

where the variance of v equals $b^2\text{var}(e)+\text{var}(w)$. The error term for $x_2$ is correlated with that of $x_1$, and the variance of $x_2$ conditional on y, is greater than the corresponding one for $x_1$ when $\text{var}(w) > (1-b^2)\text{var}(e)$. In this sense, $x_1$ may be a better indicator of y.

Anand and Harris (1991) essentially adopt the above framework in their analysis of living standards in Sri-

Lanka. Under certain conditions they are able to show that if $x_i$ is more variable than $x_j$, average savings in a given population decile will be lower when families are ranked according to $x_i$ than when they are ranked according to $x_j$. This result then allows them to perform pair-wise comparisons on the relative variability of some commonly used indicators of living standards.

The third explanation regarding the fact that alternative indicators offer different information with respect to the identification of the poor, is to be found in the assumption that the indicators are jointly correlated through their common dependence on y. This is the approach we pursue in this work. Let $g(x_i|y)$ denote the probability density function of $x_i$ when y is held fixed. We let $x_1$ and $x_2$ be related to y through the following system:

$$\begin{aligned} x_1 &= y + u_1 \\ x_2 &= \beta_2 y + u_2 \end{aligned} \qquad (3.1)$$

Define y as permanent income, $x_1$ as current income, and $u_1$ transitory income. According to Friedman (1957) p.21,

"The permanent component is to be interpreted as reflecting the effect of those factors that the unit regards as determining its capital value or wealth. ... The transitory component is to be interpreted as reflecting all 'other' factors."

We also let $x_2$ denote current consumption expenditure and $u_2$ be a disturbance term in the consumption equation. Equations (3.1) defines a pure life-cycle model, where current income and consumption are endogenous variables and are functions of permanent income. The decomposition of observed income, $x_1$, into permanent and transitory components, dates back to the work of Friedman and Kuznets (1945) on income inequality in the United States. The equation for $x_2$ was subsequently added by Friedman (1957) in his work on the consumption function. As with current income, observed consumption was decomposed into a permanent component $\beta_2 y$ and transitory element $u_2$.

The assumption that the correlation between $x_1$ and $x_2$ is solely induced by $y$ can be formally stated as follows:

$$g(x_1,x_2 \mid y) = g(x_1|y) \cdot g(x_2|y)$$

That is, when $y$ is held fixed, $x_1$ and $x_2$ must be independent. The above condition is known as the axiom of conditional independence. From the stated axiom, it follows that

$$\mathrm{cov}(u_1,u_2) = 0 \qquad (3.2)$$

Note that now the disturbances are uncorrelated, whereas when we assume $x_2$ is a function of $x_1$, the error term in the consumption equation is a function of the error term of of the income equation. In contrast to this, in the present set-up, a priori all indicators have a symmetric status with respect to $y$.

Our question is the following: can one use the information available from the two indicators, in order to predict the permanent income of the family? The answer hinges upon the identifiability of the model (3.1). The first step thus consists of examining the identification of the model relating the observed variables $X=[x_1,x_2]$, to the unobservable $y$. If we can estimate all the structural parameters of the model, than, in a second step we can attempt to construct an index of the family's permanent income.

Let us first examine the identification of model (3.1). The question of identification can alternatively be stated as follows: from the sample moments available to us, can we identify $\beta_2$, and the variance of the two error terms, say $\omega_{11}, \omega_{22}$ ?

Noting that we have three sample moments $\mathrm{var}(x_1)$, $\mathrm{var}(x_2)$, and $\mathrm{cov}(x_1,x_2)$, we can write the following identities relating sample and population moments:

$$\begin{aligned} \operatorname{var}(x_1) &= \operatorname{var}(y) + \omega_{11} \\ \operatorname{var}(x_2) &= \beta_2^2 \operatorname{var}(y) + \omega_{22} \qquad (3.3) \\ \operatorname{cov}(x_1\ x_2) &= \beta_2 \operatorname{var}(y) \end{aligned}$$

To simplify the problem further, assume $x_1$ and $x_2$ have been standardized so that $\operatorname{var}(x_1)=\operatorname{var}(x_2)=1$. We note from (3.3) that in order to identify the structural parameters of the model, we need to know the variance of y, which we will denote as $\gamma_{yy}$. Because $\gamma_{yy}$ is unknown, the model (3.1) is not identified.

Let us set $\operatorname{cov}(x_1\ x_2)$ at 0.3, the sample correlation between our income and consumption indicators. In the table below, we consider various values of $\gamma_{yy}$, and the corresponding estimates of the structural parameters. Note that in order for us to obtain any meaningful results, the variance of y has to be restricted to the interval $0 < \gamma_{yy} \leq 1$.

| $\gamma_{yy}$ | $\beta_2$ | $\omega_{11}$ | $\omega_{22}$ |
|---|---|---|---|
| 0.25 | 1.2 | 0.75 | 0.64 |
| 0.5 | 0.6 | 0.5 | 0.82 |
| 0.75 | 0.4 | 0.25 | 0.88 |
| 1 | 0.3 | 0 | 0.91 |

TABLE 1.1: Identification of model 3.1 through moment restrictions.

The introduction of prior information about $\gamma_{yy}$ may leave the practitioner uneasy, since different beliefs about the variance of the unobservable y, tend to produce different parameter estimates for the structural coefficients relating the indicators to y.

As a way out of the under-identification problem of the life-cycle model (3.1), Friedman suggested to "compute" permanent income as a weighted average of current and past values of observed income:

$$y = \frac{1}{3}\left[x_{1,t} + \frac{2}{3}x_{1,t-1} + \left(\frac{2}{3}\right)^2 x_{1,t-2} + \ldots\right] \qquad (3.4)$$

where $x_{1,t-s}$ denotes observed income s years from the present. Friedman went back 17 years in time and accordingly estimated the coefficient $\beta_2$ of the consumption equation at a value of 0.88.

Though often used in practice, the above formulation has been criticized by many researchers as being rather ad-hoc. Deaton and Muellbauer (1980) pp. 325-26 argue that "such a procedure is fraught with difficulties". One concern they raise is that:

"No account is taken of expectations about unemployment nor about the effects of current inflation rates upon anticipated future levels of income."

One thus has to rely on alternative routes to resolving the identification problem[2]. As pointed out by Goldberger (1972), the way out of the problem, is to assume the availability of another indicator of y [3]. If y is long term income status, then we can assume that say, asset holdings of the family, are also informative about y. Let $x_3$ denote this third indicator. The model (3.1), augmented by one extra equation is written below.

$$
\begin{aligned}
x_1 &= y + u_1 \\
x_2 &= \beta_2 y + u_2 \qquad (3.5) \\
x_3 &= \beta_3 y + u_3
\end{aligned}
$$

Maintaining the independence of the error terms, we can write down the moment restrictions relating the structural parameters and sample moments:

---

[2] Under some distributional assumptions, third and higher order moments may contain further information that can be used to identify the model. We therefore note that it is not always the case that with only two indicators the model is under-identified.

[3] A similar approach is followed by Van Praag et al.(1983) in the measurement of inequality, when income is subject to measurement error.

$$\mathrm{var}(x_1) = \gamma_{yy} + \omega_{11}$$
$$\mathrm{var}(x_2) = \gamma_{yy}\beta_2^2 + \omega_{22}$$
$$\mathrm{var}(x_3) = \gamma_{yy}\beta_3^2 + \omega_{33}$$
$$\mathrm{cov}(x_1, x_2) = \gamma_{yy}\beta_2$$
$$\mathrm{cov}(x_1, x_3) = \gamma_{yy}\beta_3$$
$$\mathrm{cov}(x_2, x_3) = \gamma_{yy}\beta_2\beta_3$$

Let $\sigma_{ij}$ denote $\mathrm{cov}(x_i\ x_j)$, solving the six equations above we obtain the following solutions:

$$\gamma_{yy} = \sigma_{12}.\sigma_{13} / \sigma_{23} \qquad \omega_{11} = \sigma_{11} - \sigma_{12}.\sigma_{13} / \sigma_{23}$$

$$\beta_2 = \sigma_{23} / \sigma_{13} \qquad \omega_{22} = \sigma_{22} - \sigma_{12}.\sigma_{23} / \sigma_{13}$$

$$\beta_3 = \sigma_{23.} / \sigma_{12} \qquad \omega_{33} = \sigma_{33} - \sigma_{23}.\sigma_{13} / \sigma_{12}$$

It is more convenient to set $\gamma_{yy}=1$ and to introduce a slope parameter $\beta_1$ for the equation relating the indicator $x_1$ to y. Under such normalization $\beta_i$ becomes the correlation coefficient between $x_i$ and y (and writing the moment equation for $x_1$, we see that $\beta_1$ is equal to $\gamma_{yy}$).

The general model relating the indicators X to y is written as follows:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} y + \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix}$$

or, in a more compact notation:

$$X = \beta y + U \qquad (3.7)$$

Model (3.7) is known as the *factor analysis* model. An important feature of the model is the additive structure of the covariance matrix of the observables:

$$\Sigma = \mathrm{var}(x) = \beta\beta' + \Omega \qquad (3.8)$$

The variance of X thus consists of a component $\beta\beta'$ originating from the joint dependence of the indicators on y, together with a component $\Omega$ arising from the presence of the disturbance terms.

A necessary condition for identification, is that the number of observed moments is no less than the number of free parameters (that is, the number of parameters to estimate). For the factor model (3.7), there are 2p parameters to estimate, from $p(p+1)/2$ observed moments contained in the p x p sample covariance matrix. Define d as the difference $p(p+1)/2-2p$. We have three cases to consider:

i) $d < 0$

ii) $d = 0$

iii) $d > 0$

When $d < 0$, there may exist an infinity of solutions, and the model is said to be under-identified. When $d=0$, there exist as many observed sample moments of $\Sigma$ as there are parameters to estimate, and the model is just-identified. We can therefore hope to obtain a unique set of estimates. Finally, when $d>0$, the model is over-identified. As comes out from our earlier discussion, $d \geq 0$ only when we have three indicators or more.

*Estimation*

There are two ways of dealing with the estimation of the factor analysis model. The first approach consists in specifying the distribution $g(X/y)$ of the indicators. The second approach does not assume any underlying distribution for the indicators, and selects estimates $\delta*$ from the problem of minimizing a distance $F[S,\Sigma(\delta)]$ between the sample covariance matrix S, and the matrix $\Sigma(\delta)$ induced by the model. Here we discuss the distribution-free method.

The main advantage of opting for least squares estimation methods, is that their consistency is guarantied for all distributions of X, provided the model is identified. The Generalized Least Squares estimator minimizes the distance

$$F_{GLS}\ [S, \Sigma(\delta)] = \mathrm{trace}[(S - \Sigma)V]^2$$

so that the criterion consists of minimizing the sum of squares between the various elements of S and $\Sigma$, where the distances are measured in the metric of V. When we choose V=I , we are in fact choosing estimators to minimize:

$$\Sigma\Sigma_{i,j}\ [(s_{ij} - \sigma_{ij})]^2$$

So that the choice of V=I induces an ordinary least squares method. Amongst the class of GLS estimators, the optimal choice of V (from an efficiency point of view) is a matrix which converges in probability to $\Sigma^{-1}$. For this reason the choice of $V = S^{-1}$ is immediate in practice. Conditional on the appropriate choice of V, the GLS estimator is Best Asymptotic Normal (BAN) under fairly general conditions. Sufficient conditions are that the X's are independent and identically distributed, and that fourth order moments are finite. The asymptotic covariance matrix of the GLS estimator is a factor 2/n which multiplies the inverse of the information matrix (viz the inverse of the expected value of Hessian matrix, evaluated at $\delta$*). Finally, we state an important result relating ML, GLS, and OLS. When a model is just-identified, all three estimators will produce identical solutions.

If we assume that X is a multivariate normal vector, the goodness of fit of the model can be assessed by means of a $\chi^2$ test, provided estimates are obtained by means of ML or GLS techniques. Under the null hypothesis that the model is correct, the statistic [n-1].F[ S, $\Sigma(\delta)$ ] is distributed as a $\chi^2_d$, where d is the difference between the number of sample moments, and the number of parameters that require estimation. The drawback of this test is clear: even if $\Sigma(\delta)$ were close to S, a large n would invariably lead to the rejection of the model. Another problem is that slight departures from normality, ( e.g. a distribution with some kurtosis) , render the $\chi^2$ statistic a poor approximation of the required test.

An alternative way of evaluating the fit of the model is to use the coefficient of determination CD:

$$CD = 1 - \det(\Omega) / \det(S)$$

The coefficient of determination is one way of assessing how well the indicators jointly serve the purpose of measuring the latent variables. The range of values of CD is [0,1]. A poor fit of the model is associated with a value of CD close to zero.

We have now informally discussed the identification and estimation of the model relating the indicators to the observable long term income status of the family unit(4). Other than ranking indicators in terms of their correlation with y, it is not clear how the above method might provide helpful guide-lines in identifying the poor. For example assume we estimate the model with three indicators, and say we find that the coefficient on income $\beta_1$, is estimated at 0.8, while the coefficient on consumption $\beta_2$, is equal to 0.5, and that on assets, $\beta_3$, equals 0.2. We may then decide that out of the three indicators, current income $x_1$ should be selected because of its greatest association with permanent income. The above statement is equivalent to saying that $x_1$ is the least noisy indicator of y, since the variance of $u_i$ decreases as $\beta_i$ rises.

As an alternative to selecting the least noisy indicator of y, we may wish to pool the information available from all the indicators in order to predict, in a sense, the permanent income of the families. One can then hope to rank the families in terms of their predicted permanent incomes, and separate them out between poor and

(4) Alternative expositions of the identification problem can be found in Goldberger(1972), and in Greene(1991) pp.531-35. The texts of Everitt (1984) and Bartholomew(1987) give a more detailed discussion of the question.

non-poor, according to a specified level of the poverty line.

The task of predicting y, having observed X, requires the derivation of the distribution of y conditional on X. At this stage it is necessary to make assumptions regarding the joint distribution of X and y, denoted as f(X,y). Having specified the parametric form of f(X,y), one can use Bayes' rule to arrive at h(y|X):

$$h(y|X) = f(X,y) / \int f(X,y)dy$$

In terms of the general model (3.7) relating the observed variables X to the unobservable y, we assume that $U \sim N(0,\Omega)$, and $y \sim N(0,1)$. It follows that

$$X \mid y \sim N(\beta y, \ \Omega) \qquad (3.9)$$

and $$X \sim N(0 \ ; \ \beta\beta' + \Omega)$$

Then we can use some properties of the normal distribution (e.g. see Greene(1991) p.78) to establish that the conditional distribution h( y | X) is also normal:

$$y| \ X \sim N\ [\beta' \Sigma^{-1} X \ ; \ 1- \beta'\Sigma^{-1}\beta] \qquad (3.10)$$

where $\Sigma$ was defined in (3.8) as being the variance of the vector X of indicators.

We use the mean of h(y|X) as our index of long run income, i.e.

$$E(y|X) = \beta' \ \Sigma^{-1} X \qquad (3.11)$$

The suggested index is the regression function of y conditional upon X, and is a linear function of the indicators. Before we further discuss the multiple indicator approach, we analyse our data using the suggested method. Let $\psi$ denote E(y|X). Below we refer to $\psi$ as the *multiple indicator index*.

**1.4 An Illustration**

We now illustrate the multiple indicator method on the basis of a sample of 910 families, extracted from wave XX of the U.S. based, Panel Study of Income Dynamics.

In a first stage we report results for the estimation of the model (3.7) relating three indicators, X, to an unobserved welfare standard y. Then we will construct the multiple indicator index $E(y|X)$ of (3.11), and examine the insights it offers into identifying the poor. The X variables are the following:

$x_1$ : family income to needs ratio.

$x_2$ : food expenditure/total taxable income of head and wife.

$x_3$ : total annual employment hours of head/ total annual employment+ unemployment hours of family head.

The three indicators can be taken to be correlates of permanent income. We would expect $\beta_1$ and $\beta_3$ to be positive. Even when leisure is a normal good, we expect $\beta_3$ to be positive, since the state of being unemployed is different from that of being out of the labour force altogether. We expect $\beta_2$ to be negative, since food consumption is an inferior good. We do not have other sources of information on consumption of non-durables in the PSID other than the annual household expenditure of the family on food.

The sample correlation matrix is the following:

$$S = \begin{bmatrix} 1 & & \\ -0.3005 & 1 & \\ 0.2099 & -0.2271 & 1 \end{bmatrix}$$

Parameter estimates of the model $X = \beta y + U$ are reported in table 1.2 below:

| parameter | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\omega_{11}$ | $\omega_{22}$ | $\omega_{33}$ |
|---|---|---|---|---|---|---|
| estimate | 0.527 | -0.570 | 0.398 | 0.722 | 0.675 | 0.841 |
| t-value | 9.238 | -9.475 | 8.250 | 11.857 | 10.045 | 17.197 |

Table 1.2: Parameter estimates and t-values

The coefficient of determination, which is defined as $1-\det(\hat{\Omega})/\det(S)$ takes the value 0.513.

Our coefficients are of the predicted signs. The coefficient for the employment ratio $\beta_3$ , is estimated to be about 0.4. While the coefficient of determination is not very impressive, we should note that all six parameter estimates are all highly significant.

Having estimated the structural parameters of the model, we can now derive our multiple indicator index $\psi$, on the basis of (3.11) [5]:

$$\psi = E(y|X) = 0.36x_1 - 0.41x_2 + 0.23x_3$$

We note that the budget share of food in this first application is assigned the highest weight. It is followed in importance by the income variable. The variable that carries least weight is the employment indicator of the family head.

In order to assess the impact of using $\psi$ in the identification of the poor, we construct a 4 by 4 classification matrix where families are ranked line-wise by current income $x_1$, and column-wise using the multiple indicator index $\psi$ :

---

[5] To obtain the coefficients of $\psi$, pre-multiply the inverse of the correlation matrix by the vector of estimates of $\beta$.

$$T(x_1,\psi) = \begin{bmatrix} 175 & 52 & 0 & 0 \\ 43 & 144 & 41 & 0 \\ 7 & 31 & 173 & 16 \\ 2 & 1 & 13 & 212 \end{bmatrix}$$

The main diagonal now contains 704 families (that is 77.4% of the families are ranked identically by $\psi$ and $x_1$). Using income and consumption, agreement was possible only in 445 cases (see the matrix $T(x_1,x_2)$ in section 2). In this sense, wherever we choose to set the poverty line, we are likely to find a closer ranking of families using income and $\psi$.

Let us now contrast the classification of families using consumption $x_2$, and the multiple indicator index $\psi(X)$. We rank families line-wise in order of decreasing food budget share, and column-wise in increasing order of $\psi$.

$$T(x_2,\psi) = \begin{bmatrix} 169 & 50 & 6 & 3 \\ 39 & 113 & 59 & 16 \\ 13 & 48 & 100 & 67 \\ 6 & 17 & 62 & 142 \end{bmatrix}$$

The main diagonal now contains 524 families, as opposed to the 445 families in the classification by $x_1$ and $x_2$. Let us summarize these results by means of a Venn diagram. When the poverty line is set at the bottom quartile of a given variable, each indicator will identify 227 families as being poor. As shown in figure 1.2, income, consumption, and the multiple indicator index, jointly identify 135 families as being in poverty. A further three families are commonly identified by $x_1$ and $x_2$. The multiple indicator index has in common with the income definition the 135 observations plus another 40 family units. Together with consumption, $\psi$ ranks another 34 families as being in poverty. Furthermore, we note that $\psi$ has the largest intersection with the two other sets, with a remaining 18 families non-poor according to the other two definitions.

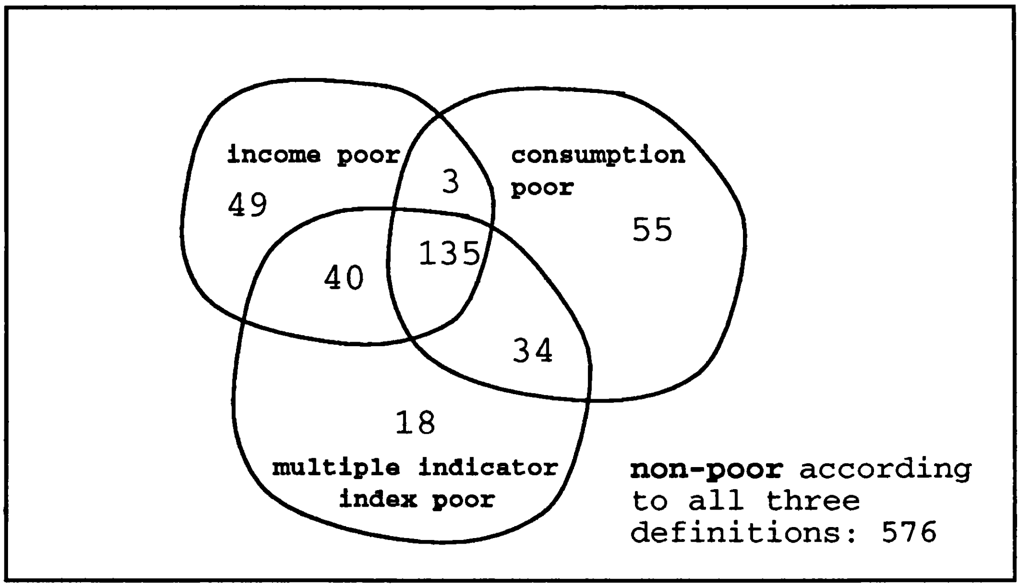


Figure 1.2: Identifying the poor using various definitions, Venn diagram

In figure 1.3 we draw a scatter diagram of the 227 families identified as being in poverty using the multiple indicator index. To avoid problems of comparison due to differing units of measurement, we plot the ranks of the observations along the income and consumption definitions. The ranks are in deviation of 227(i.e 227 is substracted from the family's rank), so that the horizontal and vertical lines divide the data into four quadrants. According to this representation, the south-west (S.W) quadrant defines the 135 families selected by the multiple indicator index who are also poor according to $x_1$ and $x_2$. Likewise, the N.E. quadrant locates the 18 families which are identified by $\psi$ as being in poverty, but not by income and consumption.

The N.W. quadrant locates the 40 families selected by $\psi$ which are income poor but not consumption poor. It is interesting to note that some of them have very low budget shares when compared to the other sample members. Turning to the 34 families in the S.E. quadrant, we note that from the selected sample, those who are consumption poor but not income poor do not have as striking features as those in the N.W. quadrant. By this we mean that these 34 families do

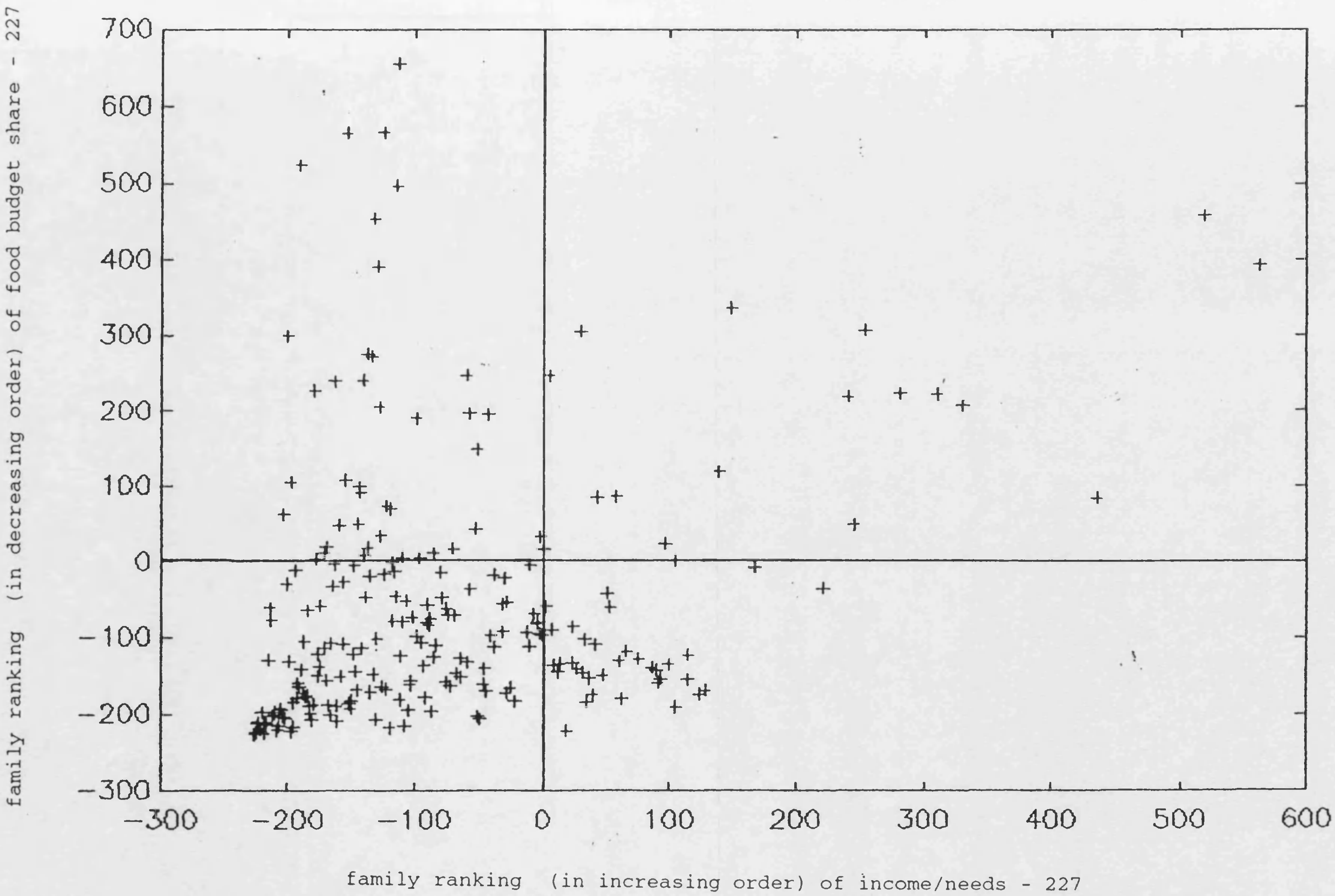


FIGURE 1.3 : Families identified as poor using the multiple indicator index: scatter plot in the income and consumption space

not have unexpectedly high income endowments, so that none of them appear to behave as outliers in the income space.

Of the remaining 18 families in the N.E. quadrant, 9 observations do not fall far from the consumption or income poverty line. There are another 9 observations, especially two in the top right corner which deserve some explanation as to their being selected poor by the multiple indicator index. These nine observations, together with the outliers in the N.W. quadrant, all share the common feature of pertaining to families where the family head experienced long spells of unemployment during 1986. This pattern is to be expected, since the employment indicator $x_3$ is used to identify the factor analysis model and to construct the multiple indicator index.

Consider now the following exercise: for various values of the poverty rate, we wish to explore the extent to which income, consumption, and the multiple indicator index $\psi$ identify the same families as being deprived. If we decide that say, the lowest ranked 10% are the poor, than we can agree on a maximum of 910/10=91 cases. In the table below $\pi$ denotes the poverty rate, and H is the head count of the poor.

| $\pi$ | H | $T(x_1,x_2)$ | $T(x_1,\psi)$ | $T(x_2,\psi)$ |
|---|---|---|---|---|
| 10% | 91 | 47 | 49 | 58 |
| 14% | 130 | 70 | 81 | 86 |
| 20% | 182 | 103 | 132 | 129 |
| 25% | 227 | 138 | 175 | 169 |

Table 1.3 : Further results based on the multiple indicator index

Table 1.3 shows that $x_1$ and $\psi$, and $x_2$ and $\psi$, will identify more families similarly (between poor versus non-poor) than the income and consumption indicators. At the lower values of the poverty rate, consumption and $\psi$ appears to coincide more often, while at the higher values of $\pi$, the pattern seems more pronounced between income and the multiple indicator index. No conclusions are to be drawn from this last statement though. These differences are too small to be of any statistical significance.

As a final summary of our results, we compute the covariance matrix C between X (first three lines) and the multiple indicator index $\psi$.

$$C = \begin{bmatrix} 1 & & & \\ -0.3005 & 1 & & \\ 0.2099 & -0.2271 & 1 & \\ 0.5270 & -0.5700 & 0.3980 & 1 \end{bmatrix}$$

As can be seen from the last line, the multiple indicator index correlates highest with all three variables in X. Note that the correlation between $\psi$ and X is nothing else than the vector $\beta$, since $E(X'\psi)=X'X(X'X)^{-1}\beta$ and $X'X$ is the sample estimate of $\Sigma$. The high degree of association between the multiple indicator index and X provides an explanation as to why the classification matrices between $\psi$ and, income and consumption respectively, offered greater scope for agreement than the analysis based on $x_1$ and $x_2$. In this sense the multiple indicator approach can be seen as a middle of the road solution between the use of income versus consumption, and vice versa.

**1.5 Assessment**

We have illustrated the multiple indicator approach above, and we have offered the view that it could be used as an intermediary solution between the exclusive use of one indicator at the expense of others. In this section we comment further on some points related to the use of the method.

Normative judgements have not been present in our discussion so far. And indeed, looking at our index $\psi$, we may be tempted to argue that it is a welfare index, based on a linear and additively separable utility function. This is certainly not the case. What $\psi$ does is to summarize the available information about the unobservable y, this latter variable is assumed to be pertinent to welfare analysis. All that our approach is aiming at is to draw on information from several indicators, as an alternative to using a unique noisy indicator variable. Note also that indicators such as race, sex, and age of the household head could be used as indicators of economic status. However, such household characteristics are not choice variables in a utility function in the way that bread and meat are.

The problem of constructing an index $\psi$ from a vector of variables X, is not a new problem in multivariate statistics. As such we have suggested to work in a factor analysis set-up (the model (3.7)). An alternative, perhaps better known technique to economists, is that of *principal component analysis* (PCA). One way to obtain the PCA model would consist in omitting the disturbance term in the system relating X to y. This would then enable us to express y as a linear combination of the x's

$$y = \delta X$$

where $\delta$ is defined as the eigen-vector corresponding to the largest eigen value of $\Sigma$[6]. In our case it is not y but $E(y|X)$ which is constructed as a linear function of X. Thus our conclusions are about expected poverty, meaning that if two families are observationally equivalent (i.e. their observed X vectors are identical), we expect them to be equally off, however due to some unobserved factors (which the disturbance term is introduced to account for), one may be better off than the other. It is ultimately up to the researcher to decide whether PCA or factor analysis is the more adequate framework to use in the context of our problem. We have opted for factor analysis, because we feel that omitting the disturbance term would be way too unrealistic. Assuming that income, consumption, and employment status, are entirely explained by a common unobservable y excluding disturbance terms, may be too restrictive a set-up.

---

(6) See Morrison (1978) pp. 267-75.

Throughout our presentation, we have assumed that the poverty line is to be set on a relative basis, so that the ranking of individuals would be informative of their economic status. We have not discussed the possibility of setting the poverty line on an absolute basis. Nothing in practice prevents us from doing so. We could fix a threshold $\psi^*$, and separate out families according to their endowments, between those who fall short of $\psi^*$, and those who do exceed the poverty line.

Two points are worth mentioning at this stage. It is often suggested that part of the task of latent variable modelling is to estimate the unobservables. It should be noted however that the framework laid out here cannot produce estimates, but at best, predictions of the extent of poverty. This is because estimating values for unobserved random variables makes strictly no sense. The second point, deals with the definition of the concept of a latent variable. It is often the case that investigators in the social sciences give names to the unobservable. The idea underlying this practice is the belief that the latent variables are existing well defined variables, and that the problem lies in their measurement. In practice however, whether we choose to define y as poverty, human capital, or permanent income, it makes no difference to our quantitative analysis. The unobservable y contains nothing more than the specification of the correlations between the indicators.

In the estimation of structural parameters, criteria such as consistency and efficiency will often guide the practitioner in his choice of estimation technique. The estimators of $\beta$ and $\Omega$ we have used are derived from sample moments. In just-identified models, the *method of moments* estimator is consistent as well as *Best Asymptotic Normal.* The method of moments estimator also presents the advantage of being distribution-free. Thus, the estimators of $\beta$ and $\Omega$ will be consistent regardless of the exact form of the population distribution of X.

We now turn to the statistical properties of the multiple indicator index $\psi$. Firstly we note that it is a linear function of the indicators. We also note that it is the *best linear predictor* of y[7]. It should be noted that this result is distribution-free. Furthermore, under the general normality assumption, $\psi$ is also the *best predictor* for y[8]. Also, when the interdependence of the indicators is fully accounted for by their common dependence on y (i.e. when the axiom of conditional independence holds), the distribution $h(y|X)$ exhausts all the information existing in the sample about y. Since X and y are both random variables, we can decompose their joint density as follows:

$$f(X,y) = h(y|X) \,.\, g(X)$$

Thus, all the information available to us about y, is contained in the conditional distribution $h(y|X)$. Any index, or predictor of y, must therefore be based on $h(y|X)$. This result is due to Bartholomew (1984), where he shows that for any $g(X|y)$ member of the exponential family of distributions, there exist sufficient summary statistics for y, which often turn out to be linear functions of the indicators.

Let us show that under the normality assumption the multiple indicator index is a sufficient statistic for y. The distribution of X when y is held fixed can be written as:

$$g(X|y) = (2\pi)^{-p/2} \, [\det(\Omega)]^{-1/2} \exp\left\{-\frac{1}{2}(X-\beta y)'\Omega^{-1}(X-\beta y)\right\}$$

and rearranging, we have

$$g(X|y) = (2\pi)^{-p/2} \, [\det(\Omega)]^{-1/2} \exp\left\{-\frac{1}{2}(X'\Omega^{-1}X)\right\} \,.\exp\left\{-\frac{1}{2}y'\beta'\Omega^{-1}\beta y\right\} \,.\exp\left\{(X'\Omega^{-1}\beta y)\right\}$$

(7) I thank Danny Quah for pointing out this result to me.

(8) See Amemiya (1985) p. 3 for a discussion regarding best linear and best predictors.

While y is a random variable, the structural parameters of the model are constants for the whole population. We may therefore define the following functions:

$$a(X) = (2\pi)^{-p/2}\ [\det(\Omega)]^{-1/2}\ \exp\left\{-\frac{1}{2}(X'\Omega^{-1}X)\right\}$$

$$b(y) = \exp\left\{-\frac{1}{2}y'\beta'\Omega^{-1}\beta y\right\} \quad \text{and} \quad d(X) = \beta'\Omega^{-1}X$$

Then $g(X|y)$ can be expressed as a member of the exponential family of distributions[9] :

$$g(X|y)=a(X).b(y).\exp[d'(X).y]$$

from which follows the sufficiency of $d(X)=\beta'\Omega^{-1}X$. Under normality, $E(y|X)=\beta'\Sigma^{-1}X = [1+\beta'\Omega^{-1}\beta]^{-1}\ \beta'\Omega^{-1}X$. In other words, $E(y|X)$ can be expressed as $m.d(X)$ with $m > 0$, so that it is also a sufficient statistic for y.

The requirement that $g(X|y)$ be a member of the exponential family for the sufficiency result to hold, is a much less restrictive condition than one would initially think. Assume X contains p variables. The exact condition for sufficiency only requires that p-1 of the x's have distributions belonging to the exponential family. Furthermore, these distributions need not be the same. $x_1|y$ could follow a gamma distribution, while $x_2|y$ could be normally distributed etc. Note also that the exponential family is broad enough to cover many of the commonly used distributions in empirical work, such as the normal, Poisson, and gamma distributions.

As an alternative to working with the multiple indicator index $\psi$, one could for instance use a weighted index of the indicators, by for example assigning equal weights to the x's:

$$\xi = (\ x_1 + x_2 + \ldots + x_p)\ /\ p$$

[9] See Mood et al.(1974) for a discussion on the exponential family of distributions and the related sufficiency results.

The weakness of an index such as $\xi$, based on arbitrary weights, is that generally it will not achieve the optimal reduction of the dimensionality of the problem ( from p to a unique dimension) in a way that no information about y is lost. Only a sufficient statistic will achieve this optimal reduction of the dimension of the data.

Finally, let us address the following question: when should we be content with using a single variable $x_i$, rather than working with the multiple indicator index? Assume we have estimated the factor analysis model (3.7), and we estimate $\beta_i$ not to be significantly different from unity, and $\omega_{ii}$ not to be significantly different from zero. Under such circumstances, we could argue that $x_i$ perfectly correlates with y, and thus that it is a suitable proxy for the unobservable.

## 1.6 Conclusions

The multiple indicator framework we have suggested is well suited to deal with a situation when a variable y pertinent to welfare analysis, is subject to measurement error. The index of y that we have proposed, is a linear function of the indicators X, and is defined as the regression function of y conditional upon X.

The multiple indicator index we have constructed from our data offered greater scope for agreement concerning the ranking of families with both consumption and income, than in the case where families were ranked using the latter two indicators. In this sense, the proposed method in this chapter can be seen as a compromise with respect to those who prefer to use consumption at the expense of income, and those who favour to work with income instead of the former.

CHAPTER 2

# The Allocation of Benefits Under Uncertainty

## 2.1 Introduction

One reason why the identification of the poor is a question of interest, resides in the fact that it enables governments to assess the likely effects of adopting various welfare policies. Poverty alleviation policies may take the form of universal population coverage, or alternatively may be restricted to certain groups of individuals. The rationale underlying the *targeting* of benefits approach, lies in the need to ensure that those in need of state help obtain relief, and conversely, that resources are not wasted on raising the welfare of the already well-off members of society.

An important assumption underlying the targeting framework, is that needs of families are observable to the decision maker. As was pointed out in the last chapter, in taking a long term perspective of poverty, we are confronted with the problem that the life-cycle/permanent incomes of individuals are generally not observable. Instead, what we may observe are imperfect correlates of long run income. In this chapter we use the concepts discussed in chapter one, in order to examine the problem of allocating benefits to families in a situation of imperfect information.

Categorical benefits are often made available to specific socio-economic groups because there exists prior information concerning their risks of being poor. If for instance there exists a rural/urban bias in the participation to education and health programmes, the government may wish to channel funds to inhabitants of rural areas, in an effort to correct the bias in the take-up of benefits. Old age and one-parent family benefits also exist because these two demographic groups are generally known to be more exposed to the risk of being in poverty. In the present chapter we investigate how information concerning the circumstances of families can be used in the decision to grant them state assistance. Rather than opting for a uni-dimensional classification related to socio-economic status (e.g. working versus jobless), we wish to use a set X of micro-data (income, expenditure, etc.) in order to decide whether a family should be granted assistance.

Assume that y is the variable underlying the welfare assessment of the family, and z is the corresponding poverty line. Poverty relief policies consist in raising y as closely as possible to z. Goodin(1985) points to the problem involved with specifying y :

"So too in social welfare policy is 'real need' an unobservable characteristic that correlates only imperfectly with standard indicators (income and wealth, family obligations and resources, etc.)"

For Besley and Coate(1992), y is defined as the "income-generating ability" of individuals. Initially Besley and Coate treat ability as an exogenous variable such as genetic endowment. At a later stage in their work, they let y be a function of the effort people put in acquiring skills, so that their ability variable approaches the concept of human capital. In this chapter, we will continue to define y as permanent income, as we have done in our previous chapter.

Assume y is the permanent income of the family and $x_1$ is its current income. One can write the following equation relating these two variables:

$$x_1 = y + \upsilon$$

where $\upsilon$ is transitory income. The question we dealt with in the previous chapter, was what could be learnt about y upon observing $x_1$. We saw then that if we observed several indicators of y ,$X=[x_1,x_2 ,\ldots,x_p]$, we could hope to identify the conditional distribution of y when X is held fixed[1]. It is not necessary to go into this problem again , so that

(1) See section 3 of chapter 1 .

in this chapter we will take it that the distribution of y conditional upon X is known to the decision maker. That is, when the decision maker observes certain characteristics of the family, he can assess its *probability* of being in need.

Consider now the problem of deciding whether to grant assistance to a family with characteristics $X_o$. The decision maker could wrongly deny it assistance, when its income $y_o$ falls short of the poverty line z. Likewise, he may decide the family qualifies for state support when in fact $y_o > z$. In order to derive the decision rule, it is necessary to make explicit the social costs of erring in either direction. It is for this reason that we believe a decision rule should be derived <u>explicitly</u> in relation to a family utility function (or a poverty measure), as well as in relation to the social opportunity costs entailed by poverty relief programmes. This will be the approach followed in the present chapter.

Two of the most studied limits on targeting are the consequences of imperfect information and the presence of claiming costs. Recognizing that identifying the poor may be problematic, Akerlof(1978) has suggested that individuals belonging to high poverty groups should "tagged" as poor for the purpose of public welfare policies. Our present analysis thus builds on Akerlof's work in that it models the probability that individuals with given characteristics face of being in poverty. Kanbur(1987) also examines the consequences of imperfect information for targeting. However, he assumes the existence only of knowledge on the poverty count of various population sub-groups, e.g. when say urban poverty is equal to 10% and rural poverty is 20%. Other researchers in the area have previously argued that the appropriate way to model the allocation of benefits under uncertainty would be to adopt a decision theoretic approach (for e.g. Goodin (1985), and Sen(1992)). The analysis presented in this chapter is a first attempt in formulating and implementing a decision-theoretic framework.

Another important contribution in the analysis of targeting under imperfect information, is the work of Besley and Coate (1992). They model the relationship between government and benefit claimants as a principal-agent model. The authors point out that the government wishes to deter the non-poor from applying for benefits, but also wishes to induce the poor in participating in welfare programmes. Besley and Coate suggest that "workfare" (i.e. work in order to qualify for state support) may be a powerful screening device in sorting out individuals with respect to their poverty status.

The decision problem considered in our work is limited to the case of granting a benefit of fixed size. Many categorical benefits in practice are of this nature. Child benefits and employment benefits in many countries operate in this fashion. In our work we do not take into account the incentive problems related to the take-up of benefits. Regarding this point, Besley(1990) writes:

"It may therefore be impossible to establish what a person would be able to earn were he or she not claiming benefits. This is a source of important incentive effects, since some individuals may choose to work less in order to qualify for benefits."

We also abstract from the costs of claiming benefits (Besley(1990)), and the more general problem of non-take-up by families who qualify for support. We also abstract from the consequences of redistributive costs associated with poverty relief programmes studied by Duclos(1992).

The plan of our chapter is the following. In section 2.2 we analyse the decision problem of granting benefits in a situation of perfect information. There we assume that y is observable to the decision maker, and we examine how changes in the poverty line, the size of the transfer, and the social opportunity cost of poverty relief policies, all affect the decision rule. In section 2.3 we analyse the decision problem in a situation of imperfect information. There we assume that y is unobserved, but the decision maker can draw evidence on the basis of a set of socio-economic characteristics of the family. In section 2.4 we give a specific example of how to construct the decision rule, and we illustrate the method with the help of the data we have used in our previous chapter. The observations from the Panel Study of Income dynamics use information on the incomes, food budget shares, and employment status of 910 American families for the year 1986. Our summary section concludes the chapter.

## 2.2 Allocation of Benefits Under Full Information

Consider a society whose judgements regarding poverty are captured by an objective function $W(u(y_1,z)\ldots,u(y_n,z))$, where W is increasing in $y_i$ if i is poor, and is non-increasing otherwise. W is additively separable in the $u_i$'s. Let there be a benefit of fixed size $\tau$, (i.e. a lump-sum transfer) available for poverty relief purposes. The decision maker has to decide to whom benefits should be granted in order to maximize post-transfer social welfare. The purpose of this section will be to analyse the problem of granting a lump-sum benefit to a family unit in a situation of perfect information. By perfect information we mean that needs or resources can be measured in a way that height and weight are, and that y is observable to the decision maker.

Assume that the utility function of a particular family takes the following form:

$$u(y,z) = \begin{cases} p(y,z) & y<z \\ k & y\geq z \end{cases} \qquad (2.1)$$

where p(.) is increasing and concave in y, and decreasing in z. We sketch the shape of the utility function below.

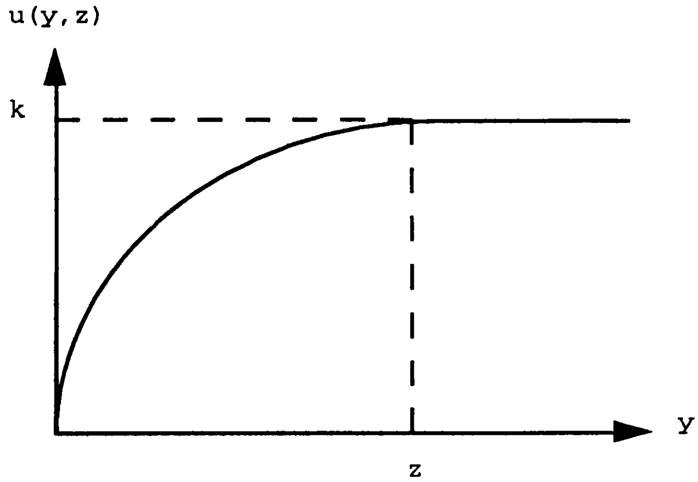


FIGURE 2.1: Family utility function

The value k is entirely arbitrary, and in what follows we set k=0.

The assumption that u(.) is non-increasing for y>z, is now common in the study of poverty (see Atkinson(1989), ch.2). It allows us to focus on the redistribution of resources to those below the poverty line. In that sense, it may seem a sensible specification in the context of poverty analysis. On the other hand, such a specification (and the associated objective of maximising the welfare of the poor) can be criticized on the grounds that it ignores the fact that redistribution to those nearing the poverty line may also be socially desirable. Atkinson (1993a) formulates his criticism concerning the "sharpness of objectives" on the grounds that:

"Such a 'sharp' representation of social objectives may not however be universally accepted. There may well be disagreement about the location of the poverty line. What one person may see as 'wasteful' expenditure on the non-poor, another may regard as contributing to the reduction of poverty....Alternatively, there may be agreement about the location of z, but concern for the 'near-poor', or the group above but close to the poverty line."

A social planner is considering granting a benefit of fixed size $\tau$ to a family with resources y. Define $\lambda$ as the social opportunity cost of government spending on welfare programmes. If the family receives assistance, its contribution to post-transfer welfare as measured in social utility units is then :

$$B(y,\tau)= u(y+\tau,z) -\lambda\tau \qquad (2.2)$$

The choices of u(.), z, and $\tau$, therefore reflect society's concern about poverty. Equation (2.2) is frequently chosen as the objective function in the economics of targeting. This formulation can be found for example in Kanbur (1987) and Duclos (1992), with the difference that here we take the transfer to be of fixed size. Though it is rarely mentioned, the formulation does away with incentive issues, in the sense that y is assumed not to be influenced by $\tau$[2] . The shadow price $\lambda$ converts government spending into social utility. It is therefore, the social opportunity cost of an extra dollar of government budget spent on poverty relief programmes.

If the family doesn't qualify for assistance, it remains at its original welfare level u(y,z), and nothing is spent by the state. The social benefit arising from the decision not to grant assistance is then :

$$B(y,0)= u(y ,z) \qquad (2.3)$$

The socially preferred decision rule is therefore

$$t = \operatorname{argmax} [ B(y,\tau) ; B(y,0) ]$$

In words, if the social welfare gain arising from the transfer outweighs its cost, the decision should be to grant support to the family:

$$\Delta B(y,\tau)= B(y,\tau) - B(y,0) \geq 0 \iff t=\tau \qquad (2.4)$$

In what follows $\Delta B(y,\tau)$ is defined as the *net social benefit* arising from the transfer.

---

[2] Regarding this point, see Besley(1990), and Besley and Coate(1992).

Let us define $y_c$ as the value of y that equates $B(y,\tau)$ to $B(y,0)$. The critical (permanent) income $y_c$, is the income above which the social cost of the transfer exceeds the corresponding benefit. The critical value $y_c$ is defined via the following equality :

$$u(y+\tau,z) - u(y,z) = \lambda\tau$$

or, in more compact notation,

$$\Delta u(y,\tau,z) = \lambda\tau \qquad (2.5)$$

$\lambda$ proxies the Lagrange multiplier on the government's budget constraint, and $\Delta u(y,\tau,z)$ is the welfare gain to the family induced by the transfer. Using the Implicit Function theorem , we can express $y_c$ as a function of $\lambda, \tau$, and z.

$$y_c = \xi(\lambda,\tau,z) \qquad (2.6)$$

$y_c$ is decreasing in $\lambda$, meaning that fewer people may qualify for state assistance if the social opportunity cost of such programmes is on the increase. The critical income level is increasing in z, if the cross derivative of $p(y,z)$ with respect to y and z is positive. For example, if $p(y,z)$ is the utility function:

$$p(y,z) = -[z-y]^a/a \text{ and } a>1 \qquad (2.7)$$

then $y_c$ is increasing in z when $a>1$, meaning that raising the level of the poverty line will allow more people to qualify for state support. On the other hand, if for example we follow Watts (1968) in specifying $p(y,z)=\log(y/z)$, the cross derivative of $p(y,z)$ will be null. In the case of the Watts measure, the critical income $y_c$ will be independent of where we choose to set the poverty line.

Finally, we note that increases in $\tau$, the size of the government transfer, will go in the direction of reducing the critical income $y_c$ . This is because at $y_c$, the marginal social benefit arising from the transfer, $u'(y_c+\tau,z)$, is lower than $\lambda$, the corresponding marginal social cost. We illustrate this point with the help of the diagram drawn in figure 2.2 below.

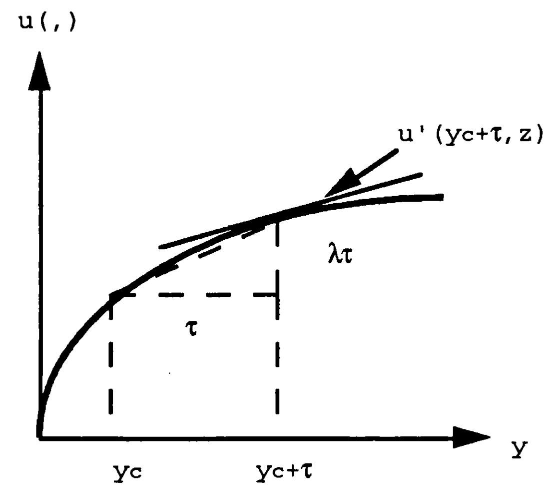


FIGURE 2.2: Government transfers and the critical income

Given equation (2.5), the difference between pre-transfer and post-transfer utility levels at $y_c$ must be $\lambda\tau$. The slope of the dashed line joining $u(y_c,z)$ to $u(y_c+\tau,z)$ is therefore $\lambda\tau / \tau = \lambda$. Because the marginal utility of income is a decreasing function, from the Mean Value Theorem we know that there exists an income level $y_m$ such that $u'(y_m,z)=\lambda$, and $y_c < y_m < y_c+\tau$. The concavity of $u(,)$ in $y$ implies that $u'(y_m,z) > u'(y_c+\tau,z)$, so that $\lambda > u'_y(y_c+\tau,z)$. Thus, $y_c$ is a decreasing function of $\tau$. In other words, if the state wishes to be more generous in terms of the support it grants to families, the resulting effect will be a reduction in the extent of population coverage[3] .

Let us illustrate these results with the help of the function $p(y,z)$ of (2.7). Consider a family with income $y=z-\varepsilon$, where $\varepsilon \geq 0$. For the family to qualify for assistance, it must be that:

$$u(z - \varepsilon + \tau, z) - u(z - \varepsilon, z) \geq \lambda\tau$$

---

(3) For a proof of the above statements see the appendix.

Given the function p(y,z) of (2.7), the first term of the left hand side of the inequality cancels out to zero when $\varepsilon < \tau$. Conditional on certain parameter restrictions we can therefore explicitly solve for $\varepsilon$ as follows:

$$[\varepsilon]^a/a \geq \lambda\tau$$

The family qualifies for benefits when

$$\tau > \varepsilon \geq [a\lambda\tau]^{1/a} \qquad (2.9)$$

A necessary condition for (2.9) to hold is that

$$\tau > [a\lambda]^{1/a-1} \qquad (2.10)$$

That is, conditional on (2.9), we can express the critical income as being

$$y_c = z - [a\lambda\tau]^{1/a}$$

A family with income y then qualifies for state support if $y \leq y_c$. Increases in z will therefore raise the critical income $y_c$. Also, in the neighborhood of (2.9), increases in $\lambda$ and $\tau$ will induce a reduction in the value of the critical income.

### 2.3 Consequences of Imperfect Information

Allocating benefits when uncertainty prevails about who the poor are, raises more complications. The problem can be (unfairly) summarized by noting that two types of errors may occur in the decision making process. When a poor person is denied a benefit (t=0) a type I error occurs. When a non-poor is granted assistance, a type II error results. Many economists (presumably) believing that type II errors occur too often in practice, have suggested that benefits should only be restricted to certain groups. The motivation behind *efficient targeting* was to avoid the extensive wastage originating from type II errors.

| decision | state of | nature |
| --- | --- | --- |
| | poor | non-poor |
| t=0 | type-I error | 0 |
| t=τ | – | type-II error |

TABLE 2.1: Decisions under uncertainty, types I and II errors

For instance see the work of Glewwe (1990). Others, such as Atkinson (1993a) and Sen (1992), have expressed scepticism with regards to targeting and the concept of efficiency underlying it. Note that in a decision theoretic framework, the argument for universal benefits can be made on the basis that welfare costs of type I errors may be very high, despite the fact that type II errors are more frequent[4].

The difference between the present decision problem and that of the full information case lies in the fact that y is not observed by the decision maker. The social benefit that results from a transfer to a person is thus a random variable. When $y \geq z$, the benefit is zero since the applicant is non-poor. Estimating the distribution of y when X is held fixed, is then the key element of an expected utility maximizing decision maker. His cost /benefit exercise of granting assistance to an applicant will be carried out by means of the comparison of the social cost , and the expected value of the social benefit of the transfer.

In analysing the decision problem under uncertainty we assume that y is not observed, but that it is related to a set X of observed indicators through the following relation:

---

(4) For a brief introduction to statistical decision theory, see Barnett(1982) ch.7.

$$y=\alpha'X + e \qquad (3.1)$$

Define $\delta$ as the set of structural parameters relating the random variable y to the random vector X and the error term e. In a first stage we will assume that the distribution of y conditional upon X, $h(y|X;\delta)$, is known to the decision maker, though y is not observed. In sub-section 2.3A below, we extend the discussion of the previous section to analyse the decision making process when y is unobserved, but when $h(y|X;\delta)$ is known . Then in sub-section 2.3B, we draw on the results of our previous chapter, in order to carry the discussion through to the case when the conditional distribution of y upon X is unknown.

*2.3A $h(y|X;\delta)$ is known*

If y is not observable to the decision maker, it is legitimate to question how the conditional distribution $h(y|X;\delta)$ may be known. Glewwe(1990) suggests that $h(y|X;\delta)$ may be inferred on the basis of household surveys, provided these constitute a source of accurate information:

"First, the income (or expenditure) data must be accurate, otherwise it will introduce a source of error and make it difficult to judge the accuracy of targeting".

Typically, due to the way the tax/benefit system operates, the government cannot ascertain the exact income of a family. Nonetheless, it possesses more accurate knowledge about the overall distribution of income. In the framework of Glewwe, y is a variable which is measurable in money terms, so that his assumption that consumer surveys provide information on y becomes a valid one[5]. The reader may then wish to inquire as to why the decision maker cannot observe y, while this information is known to the benefit applicant.

[5] When y is permanent income, as we choose to define it in our work, the measurement error problem is bound to remain in surveys based on cross-section type sampling.

Glewwe's explanation is an account of the often discussed adverse selection problem related to the take-up of benefits:

"Given that both governments and non-governmental organizations have limited resources, it is important that assistance is not mistakenly given to the non-poor, who may attempt to gain access to benefits by misrepresenting their income status."

When $h(y|X;\delta)$ is known but y is unobserved, the net social benefit arising from the transfer also becomes a random variable:

$$\Delta B(y,\tau) = \begin{cases} -\lambda\tau & \text{if the family is non-poor} \\ \Delta u(y,\tau,z)-\lambda\tau & \text{otherwise} \end{cases} \qquad (3.2)$$

The problem of allocating benefits under uncertainty is then one of choosing a decision rule t(X), so as to maximize social welfare.

| decision | state of nature | |
|---|---|---|
| | poor | non-poor |
| $t=0$ | 0 | 0 |
| $t=\tau$ | $\Delta u(y,\tau,z)-\lambda\tau$ | $-\lambda\tau$ |

TABLE 2.2: States of nature, decisions, and social payoffs

The net benefit arising from the transfer, from society's point of view, therefore depends on the decision t(X). Since the expected net social benefit under t=0 is null, a consistent decision rule with respect to expected utility maximization is to grant the applicant assistance if the expected value of the net social benefit from the transfer is positive.

*Example*

Let us illustrate the above ideas by means of an example. Assume the utility function takes the form:

$$u(y,z) = \begin{cases} -c[z-y] & y<z \\ 0 & y\geq z \end{cases}$$

The state is considering granting an infinitesimal transfer of size $\tau$ to families. A family should then qualify for state assistance if the expected marginal benefit from the transfer outweighs the marginal cost, i.e. if :

$$\frac{\partial}{\partial\tau}\int_{-\infty}^{z-\tau} -c\ [\ z - y - \tau\ ]\ f(y)\ dy \ \geq\ \lambda$$

Furthermore, assume $y=x+\varepsilon$, where $\varepsilon$ is an error term with a normal $N(0,\omega)$ distribution. The decision maker observes x but not y. Then,

$$\text{Prob}[y < z-\tau] = \text{prob}[\varepsilon < v(\tau)] \text{ with } v = z-x-\tau.$$

Likewise, $-c[z-y-\tau] = -c[v(\tau)-\varepsilon]$ so that

$$\int_{-\infty}^{z-\tau} -c\ [\ z - y - \tau\ ]\ f(y)\ dy \ = \int_{-\infty}^{v(\tau)} -c\ [v(\tau) - \varepsilon]\ f(\varepsilon)\ d\varepsilon$$

and from Leibniz' rule, it follows that

$$\frac{\partial}{\partial\tau}\int_{-\infty}^{v(\tau)} -c\ [v(\tau) - \varepsilon] f(\varepsilon)\ d\varepsilon \ = -c\ \frac{dv}{d\tau}\frac{\partial}{\partial v}\left\{\int_{-\infty}^{v(\tau)} [v(\tau) - \varepsilon] f(\varepsilon)\ d\varepsilon\right\}$$

$$= c\ \frac{\partial}{\partial v}\int_{-\infty}^{v(\tau)} [\ v - \varepsilon\ ]\ f(\varepsilon)\ d\varepsilon$$

Upon integrating, we get

$$\int_{-\infty}^{v(\tau)} [\ v - \varepsilon\ ]\ f(\varepsilon)\ d\varepsilon = v\ F(v)\ -\ v\ F(v)\ -\ \int F(\varepsilon) d(\varepsilon)$$

so that $$\frac{\partial}{\partial\tau} \int_{-\infty}^{v(\tau)} -c\ [v(\tau)\ -\ \varepsilon]\ f(\varepsilon)\ d\varepsilon\ = c\ F(v).$$

A family is therefore granted a benefit if $c\ F(z-x-\tau) \geq \lambda$. Assume that $c = 2\lambda$. We would then grant assistance to all families who have endowments x such that $x + \tau \leq z$.

* * *

Let $E(\Delta B)$ denote the expected net benefit from the transfer. The expected benefit criterion is defined as follows:

$$E(\Delta B) = E[\Delta u(y,\tau,z)] - \lambda\tau \qquad (3.3)$$

Now consider a family which is just at the margin of qualifying for state support when the poverty line is $z_o$, the size of the benefit equals $\tau_o$, and the social opportunity cost of government transfers is $\lambda_o$. For this family $E(\Delta B)=0$, so that

$$E(u(y+\tau_o, z_o)] = E[u(y, z_o)] + \lambda_o\tau_o \qquad (3.4)$$

Differentiating (3.3) at the critical value $E(\Delta B)=0$, we find that increases in $\lambda$ and $\tau$ will reduce the expected value of the social benefit arising from the transfer. These points are illustrated in figure 2.3, and derived in the appendix of the chapter.

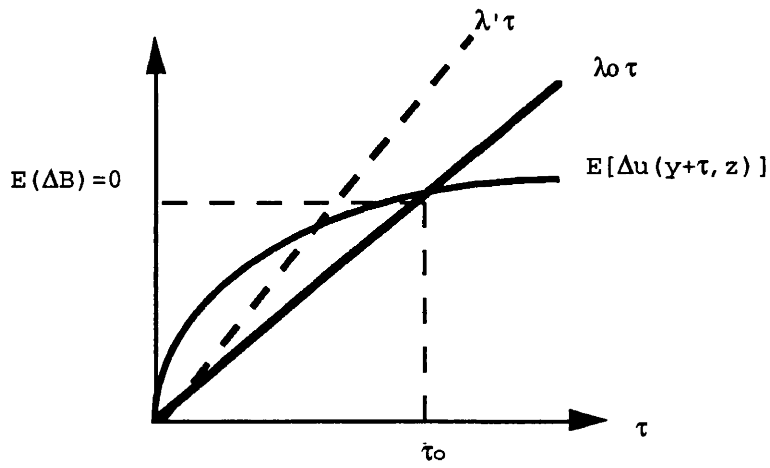


FIGURE 2.3: Changes in the social opportunity cost of poverty relief policies

Increases in $\lambda$ will shift the line $\lambda_0\tau$ upward to $\lambda'\tau$, so that $E(\Delta B)$ falls. If $\tau$ is increased beyond $\tau_0$, the expected benefit from the transfer increases less rapidly than the corresponding social cost, resulting in a decrease in $E(\Delta B)$. Finally, the expected benefit from the transfer is increasing in z when the cross derivative of u(,) is positive (and conversely when the cross-derivative is negative). These results echo the findings of the full information set-up.

*2.3B $h(y|X;\delta)$ is unknown*

If we believe that the state should give assistance to families in need, for all practical purposes we have to be able to define the concept of needs. Many economists are aware of the problems involved in assessing needs and living standards. Income and consumption with their many limitations have been frequently chosen as the variables forming the basis of judgement with regards to needs. Chaudhuri and Ravallion (1992) have argued that annual incomes were subject to substantial variability, and that many households that were not poor for a particular year, had little chance of escaping poverty in the long run. If we were to adopt long term income as the criterion for assessing needs, we must recognize that cross-section data cannot provide direct information on permanent income[6]. It is therefore worth inquiring what can be inferred about y, when we observe cross-section data X such as income and consumption, but y is not observed.

We saw in the previous chapter that under certain conditions[7], if we postulated a *factor analysis* model between X and y ;

$$X = \beta y + U \qquad (3.3)$$

we could estimate the vector $\delta$ of structural parameters of the model (3.3). The set $\delta$ contains the vector $\beta$, and the covariance matrix $\Omega$ of the error vector U.

One of the ingredients of expected utility analysis is the subjective probabilities that a decision maker attaches to the occurrence of particular events. If the relevant variable is continuous, these beliefs will be defined via a continuous probability distribution. If we postulate that X and y have a joint probability distribution $g(X,y;\delta)$, then it follows from Bayes' rule that

$$h(y|X;\delta) = g(X,y;\delta) / \int g(X,y;\delta)dy$$

Therefore, equipped with consistent estimators of $\delta$, we can estimate the distribution of y conditional upon X. The estimator of $\delta$, we note, is derived without observations on y, and is distribution-free. The choice of the parametric relation between X and y is on the other hand arbitrary, and normality is often chosen for convenience purposes. Prior knowledge on the *distribution* of life-time incomes is to say the least, scarce. Nonetheless, there exist human capital theories of life-cycle earnings, which may prove useful as a starting point in elaborating on the distribution of

---

(6) Panel data , as we saw in our previous chapter, can also be used to predict permanent income. However, one cannot expect a welfare officer to turn down young benefit applicants, because say, they have only worked for one year, and that a unique observation does not provide sufficient information about long term income.

(7) These are essentially statistical identification conditions.

permanent incomes. (See Weiss(1985) for a survey on the determinants of life cycle earnings.)

**2.4 An example**

In this section we construct an example to illustrate the problem of allocating benefits under uncertainty. We then use the data analyzed in the previous chapter, to examine how changes in the poverty line z, the size of the transfer $\tau$, and the social opportunity cost of poverty alleviation $\lambda$, affect the number of families which may qualify for benefits. In a first stage, we develop the expected net social benefit criterion using a specific utility function. This task will be undertaken in sub-section 2.4A. In the second sub-section, we analyze the characteristics of the families which qualify for state support, using the decision rule derived in sub-section 2.4A.

*2.4A Specification of the Expected Net Social Benefit Function.*

Our starting point is to specify a form for the utility function (2.1). In section 2.2, we stated that the function

$$u(y,z) = \begin{cases} -[z-y]^a/a & a>1 \quad \text{and} \quad y<z \\ \\ 0 & y\geq z \end{cases} \qquad (4.1)$$

satisfied the requirements that

u1)- u is non-increasing in z and non-decreasing in y.

u2)- for y<z u(.) is increasing and concave in y.

u3)- for y≥z u is constant in y.

u4)- The cross-derivative of u is non-negative.

Under conditions u1-u4, we established that the critical income level $y_C$, and the expected net benefit from the transfer, $E(\Delta B)$, were decreasing in $\lambda$ and $\tau$, and increasing in z. We therefore retain the utility function (4.1), where we set a=2.

The decision maker grants a benefit to the family if the expected social benefit from the transfer outweighs its cost, i.e. if

$$E(\Delta B) = \int_{-\infty}^{z-\tau} u(y+\tau, z)\; h(y|X,\delta)\,dy \; - \int_{-\infty}^{z} u(y,z)\; h(y|X,\delta)\,dy \geq \lambda\tau \qquad (4.2)$$

Let $\mu(x)$ and $\sigma$ respectively denote the expected value, and standard deviation, of the distribution of y when X is held fixed. Define the random variable $\eta$ as follows:

$$\eta = [y - \mu(x)]/\sigma \qquad (4.3)$$

Then, $\eta$ has zero mean and unit variance. Likewise, define the following variables:

$$w = [z - \tau - \mu(x)]/\sigma \text{ and,} \qquad (4.4a)$$

$$v = [z - \mu(x)]/\sigma \qquad (4.4b)$$

It follows from the definitions of w and v that

$$\text{Prob}(y<z) = \text{Prob}(\eta<v) \text{ and}$$

$$\text{Prob}((y+\tau<z) = \text{Prob}(\eta<w)$$

Also, we have from (4.4a) that $(z-y) = \sigma(v-\eta)$, so that

$$\text{for } y<z,\ u(y,z) = -\sigma^2(v-\eta)^2/2 \qquad (4.5)$$

Thus substituting (4.5) into (4.2), we obtain the following expression for the expected net social benefit criterion:

$$E(\Delta B) = \frac{\sigma^2}{2}\left\{\int_{-\infty}^{v} (v-\eta)^2\ h(\eta)\,d\eta - \int_{-\infty}^{w} (w-\eta)^2\ h(\eta)\,d\eta\right\} - \lambda\tau \qquad (4.6)$$

What remains for us to do in order to complete the task of specifying the expected benefit function, is to choose a parametric form for the joint distribution of X and y. In section 1.3 of our previous chapter, we saw that under the joint normality assumption, the distribution of $y|X$ was also normal with mean :

$$\mu(x) = \beta'\,(\beta\beta' + \Omega)^{-1}\,X\ ; \qquad (4.7a)$$

and variance $\sigma^2 = \{1 - \beta'\,(\beta\beta' + \Omega)^{-1}\,\beta\}$ (4.7b)

so that from (4.3),

$$\eta = [y - \beta'(\beta\beta' + \Omega)^{-1}\,X]/\{1 - \beta'\,(\beta\beta' + \Omega)^{-1}\,\beta\}^{1/2} \qquad (4.8)$$

and, $h(\eta) = (2\pi)^{-1/2}.\exp(-\eta^2/2)$ (4.9)

Substituting (4.8) and (4.9) into (4.6) completes the specification of the expected benefit function, as well as the decision rule. If $E(\Delta B) \geq 0$ the decision rule dictates that a family ought to be granted assistance. In the sub-section below , we report some results on the decision to grant state support to families when z, $\lambda$, and $\tau$, are allowed to vary.

*2.4B Results*

We are here making use of the data we have looked at in our previous chapter. For the year 1986, we have information on 910 families from the Panel Study of Income Dynamics, from which we draw evidence from the following variables:

$x_1$ : family income to needs ratio.

$x_2$ : food expenditure/total taxable income of head and wife.

$x_3$ : total annual employment hours of head/ total annual employment+unemployment hours of head.

From the results reported in table 1.2 of chapter 1, and under the general normality assumption, we can derive the following conditional distribution for y:

$$y|X \sim N(\ 0.36x_1 - 0.41x_2 + 0.23x_3\ ;\ 0.513)$$

In our first application, we set the poverty line z at a level such that 10% of the population is expected to be poor. By this, we mean that the inequality $0.36x_1 - 0.41x_2 + 0.23x_3 < z$, holds for 10% of the families included in our sample.

| $\pi$=10% | $\tau$=0.1 | $\tau$=0.2 |
|---|---|---|
| $\lambda$=0.5 | 70 | 66 |
| $\lambda$=1 | 31 | 28 |
| $\lambda$=1.5 | 16 | 14 |

**Table 2.3: Allocation of benefits for a poverty head-count of 10%**

Table 2.3 above reports the number of families which ought to be granted benefits for various values of $\lambda$ and $\tau$. Since the income to needs ratio ($x_1$) is equal to unity for a family who is at the U.S. official poverty line, we can interpret $\tau$ as being a lump-sum transfer adjusted for family size. Under this interpretation, a value $\tau$=0.1 is taken to signify a transfer amounting to 10% of the family's "official" needs. When $\lambda$=1, we are in a situation whereby a dollar spent on poverty alleviation has a social cost of unity. On the other hand, when $\lambda$=1.5, one dollar spent on poverty alleviation is equivalent, in social accounting units , to a reduction of spending on alternative government programmes by 1.5 dollars, etc.

The findings are in conformity with what follows from the discussion of sections 2.2 and 2.3. Increase in $\lambda$ and $\tau$ reduce the number of families which may be granted the benefit. For instance, at $\tau=0.1$ and $\lambda=0.5$, 70 families qualify for state support, and 66 families when the size of the transfer is doubled and $\lambda$ remains fixed at 0.5. At the other end, for $\lambda=1.5$, these figures drop to 16 and 14 families respectively.

Let us for now examine the 14 most deprived families. These are the ones that qualify for state help when $\lambda=1.5$ and $\tau=0.2$. Below we refer to this sub-sample as S14, and to the entire sample as S910. If for a particular family in 1986, the family head has not been unemployed, the family spends less than one third of its income on food, and its income is at least as high as its needs, we will refer to the family as being in *acceptable economic conditions*. If the family violates n out of the three conditions, we will say that it is poor on n indicators. For instance, if the family spends half of its income on food, the head is never out of a job, and the family crosses the poverty line in the income space, according to this multiple deprivations exercise, the family is poor on the basis of one indicator. The main features of sub-sample S14 are summarized below:

-One family is poor on the basis of one indicator, 11 on the basis of 2 indicators, and two on the basis of all three indicators. Thus, none of these fourteen families is living in *acceptable economic conditions*.

-Only three family heads did not experience unemployment spells. Only one family spends less than a third of its income on food. Five out of these families are income poor ( there are seven income poor families in the entire sample).

-The average family in this sub-sample has 0.17 times the average income of the entire sample. For S14, the food budget share is 10 times the average for S910, and the employment ratio is half that of the corresponding S910 average.

In table 2.4 we repeat the above exercise, with the difference that we raise the poverty line, to obtain a poverty head-count equal to 20%. Because we chose a utility function with a non-negative cross derivative with respect to y and z, we established that the critical value $y_c$, ( and thus the number of families which qualify for assistance) was non-decreasing in z. This is why, in table 2.4, when the poverty line is raised, the number of families which qualify for benefits are never fewer. The degree of coverage varies from 86 families for $\lambda$=0.5 and $\tau$=0.1, to 16 families for $\lambda$=1.5. These figures all increase, with the exception of the case where $\lambda$=1.5 and $\tau$=0.1. There , the number remains at 16 families. We now summarize the characteristics of the 33 families which qualify for support when $\lambda$=1 and $\tau$=0.1. We denote this sub-sample as S33.

| $\pi$=20% | $\tau$=0.1 | $\tau$=0.2 |
|---|---|---|
| $\lambda$=0.5 | 86 | 80 |
| $\lambda$=1 | 31 | 28 |
| $\lambda$=1.5 | 16 | 14 |

Table 2.4: Allocation of benefits for a poverty head-count of 20%

-Six out of the 33 family heads did not experience unemployment spells.

-Six families have a food budget share lower than 1/3.

-All seven income poor families of the entire sample S910, are present in the sub-sample.

-Three families are poor on the basis of all three indicators. Twenty are poor on two out of the three indicators, and 8 are poor on only one indicator. None of these families are therefore living in acceptable economic conditions.

-For S33, the average income is a quarter of the average income of the entire sample. The food budget share is 6 times the average for S910, and the employment ratio is a fraction 0.57 of the corresponding S910 average.

It should be noted that the framework we have suggested for the allocation of benefits under uncertainty is compatible with a situation where the optimal poverty alleviation policy is to extend coverage to the entire population. It follows from our discussion that this scenario is more likely the higher the poverty line is, and the lower the size of the benefit and social opportunity costs of government transfers are.

| $\pi$=25% | $\tau$=0.01 | $\tau$=0.02 |
|---|---|---|
| $\lambda$=0.01 | 892 | 892 |
| $\lambda$=0.02 | 875 | 875 |

Table 2.5: Approaching universal population coverage

In table 2.5 we consider various values of $\lambda$ and $\tau$ which would nearly entail universal population coverage when the poverty head-count is 25%. Since the sample consists of 910 families, we note for instance that for a benefit size of 0.01, a value 0.01 for $\lambda$ ensures that 98% of sample families qualify for state assistance.

To summarize our findings, we compare S14, S33, and the sample S910 of 910 families. In the table below, "o" refers to the families which are in acceptable economic conditions. z denotes families which are poor on one indicator, zz for those who are poor on the basis of two indicators etc. Finally, the last three columns contain the means of the three indicators, expressed as ratios with respect to the corresponding averages for the entire sample of 910 families.

| Sample $S_i$ | o | z | zz | zzz | mean of $x_1$ | mean of $x_2$ | mean of $x_3$ |
|---|---|---|---|---|---|---|---|
| $S_{14}$ | 0 | 1 | 11 | 2 | 0.17 | 10.11 | 0.53 |
| $S_{33}$ | 0 | 8 | 22 | 3 | 0.23 | 6.03 | 0.57 |
| $S_{910}$ | 717 | 155 | 35 | 3 | 1 | 1 | 1 |

Table 2.6: Characteristics of recipients for various degrees of coverage

The last line is of interest since it defines the characteristics of recipients in a situation where the government introduces a universal benefit system. Let us compare this bench-mark assumption with the case where coverage is restricted to 14 families. The families of S14 have the lowest average income endowments, 10 times the average food budget share, and half the employment ratio of the average family head of S910. When the parameters of the economy are altered so as to increase coverage to 33 families, the sample S33 is less poorly endowed than S14, but still significantly worse off than the sample S910. We can summarize our findings by stating that as we move from restricted coverage towards a universal benefit system, the pool of recipients reveal on average better income endowments , more favourable employment ratios , and lower food budget shares. In this sense, we feel that the framework suggested for the allocation of benefits in a situation of uncertainty, seems adequate.

### 2.5 Summary

The purpose of this chapter was to analyze the problem of benefit allocation in a situation of imperfect information. We have suggested the use of socio-economic indicators as a means of assessing the probability that a family is in poverty. The decision rule then derived consisted in granting state assistance to a family if the expected value of the social benefit from the transfer outweighed the corresponding cost. This rule arose from a cost-benefit analysis problem in an environment of uncertainty. Alternatively, it can be viewed as the solution to a problem in statistical decision theory.

In order to keep the framework tractable at the empirical level, we have made some important omissions. Incentive issues, the problem of take-up, and redistributive costs, have all been discarded from the analysis. In order to improve in the direction of making our work relevant to a real world situation, the various omitted factors need to be accounted for in the cost/benefit exercise of granting state assistance.

## 2.A Appendix

In this appendix we prove two propositions. The first of these pertains to the critical income $y_c$ (section 2.2), and the second, to the expected benefit criterion (section 2.3).

***Proposition 1:*** *The critical income $y_c$ is a decreasing function of $\lambda$ and $\tau$. It is an increasing function of z if $u''_{yz}(y,z) > 0$, decreasing if $u''_{yz}(y,z) < 0$, and independent of z if $u''_{yz}(y,z) = 0$.*

Sufficient conditions for the use of the implicit functions theorem are given in Chiang (1984) p. 206. In the context of our work, they can be stated as follows:

(i) at a point ( $y_c$, $z_o,\lambda_o,\tau_o$) satisfying the condition

$$u(y_c+\tau_o\ ,z_o)\ -\ u(y_c\ ,z_o)\ =\ \lambda_o\tau_o \qquad (a.1)$$

$\Delta B$ has continuous partial derivatives with respect to $y,z,\lambda$, and $\tau$.

(ii) At the point $(y_c,z_o,\lambda_o,\tau_o)$, the partial derivative of $\Delta B$ with respect to y is different from zero.

Condition (i) is always satisfied provided $y_c < z$. From (a.1) it follows that $y_c < z$, since if $y_c \geq z$ it must be that $\Delta B=-\lambda\tau \neq 0$.

Differentiating $\Delta B$ with respect to y at $(y_c, z_o, \lambda_o, \tau_o)$, we have

$$\Delta B'_y = u'_y(y_c+\tau_o, z_o) - u'_y(y_c, z_o) < 0$$

since $y_c < z$. Condition (ii) is therefore satisfied. There thus exists a neighborhood of $(y_c, z_o, \lambda_o, \tau_o)$ where the implicit function $y_c = \xi(\lambda, \tau, z)$ is defined.

Totally differentiating (a.1) holding z and $\tau$ constant, we have

$$[\, u'_y(y_c+\tau, z) - u'_y(y_c, z)\,]dy_c = \tau\, d\lambda$$

$y_c$ is therefore decreasing in $\lambda$. Likewise,

$$[\, u'_y(y_c+\tau_0, z) - u'_y(y_c, z)\,]\, dy_c = [\lambda_0 - u'_\tau(y_c+\tau_0, z)]$$
$d\tau$

From the Mean Value Theorem there exists $y_m$ such that $u'_y(y_m, z)=\lambda_0$, and $y_c < y_m < y_c+\tau_0$. The concavity of u(,) in y implies that $u'_y(y_m, z) > u'_y(y_c+\tau_0, z)$, so that $\lambda_0 > u'_y(y_c+\tau_0, z)$ (see figure 2.2). It follows that $y_c$ is also decreasing in $\tau$.

Totally differentiating (a.1) holding $\lambda$ and $\tau$ constant, we have

$$[u'_y(y_c+\tau, z) - u'_y(y_c, z)]dy_c + [u'_z(y_c+\tau, z) - u'_z(y_c, z)]dz = 0$$

The second bracketed expression is positive when $u''_{yz} > 0$. Thus $dy_c/dz > 0$ if $u''_{yz} > 0$, etc.

***Proposition 2:*** *At $E(\Delta B)=0$, the expected value of the net social benefit from the transfer is a decreasing function of $\lambda$ and $\tau$. It is increasing in z if $u''_{yz}(y,z) \geq 0$, and decreasing if $u''_{yz}(y,z) < 0$.*

At the critical value $E(\Delta B)=0$, the following equality holds:

$$E[u(y+\tau_o, z_o)] - E[u(y, z_o)] = \lambda_o \tau_o \qquad \text{(a.2)}$$

Differentiating $E(\Delta B)$ with respect to $z, \lambda$, and $\tau$, we have:

$$E(\Delta B'_z) = E[u'_z(y+\tau_o, z) - u'_z(y, z)] > 0 \text{ if } u''_{yz}(y, z) > 0 \text{ , etc.}$$

$$E(\Delta B'_\lambda) = -\tau_o < 0$$

also, $$E(\Delta B'_\tau) = E[u'_\tau(y+\tau_o, z)] - \lambda_o$$

From the Mean Value Theorem, there exists $\tau_m$ such that $E[u'_\tau(y + \tau_m, z)] = \lambda_0$ and $0 < \tau_m < \tau_0$. The concavity of $u(,)$ in $y$ implies that $E[u'_\tau(y + \tau_o, z)] < \lambda_0$ (see figure 2.3). Therefore, at the critical value $E(\Delta B) = 0$, increases in $\tau$ will reduce the expected value of the net social benefit arising from the transfer.

CHAPTER 3

## Poverty and the Economics of Information

### 3.1 Introduction

Chapters one and two have discussed questions related to the identification and allocation of benefits to the poor. This chapter differs from the previous two in the sense that the discussion is centred around examining specific economic mechanisms which cause poverty to arise, as opposed to treating deprivation as a given, or exogenous, phenomenon.

The Public Finance literature is rich with examples in which the market allocation of resources is not Pareto efficient. Missing markets, imperfect information, rationing, and other obstacles to trade, will typically result in a negation of the optimality property of the competitive equilibrium. Less work though has been done in the direction of assessing the impact of market imperfections on the incidence of poverty. An often discussed example in the area, is the degree to which information asymmetries in the credit market distort the allocation of investment from the social optimum. In this chapter we wish to extend the discussion in the direction of examining the likely effects of credit market imperfections on the level of poverty. One aim of the work is to show that credit market imperfections need not have an adverse effect on the level of poverty. Furthermore, we wish to explain how an economy operating subject to an imperfect credit market, may attain a lower level of poverty than would be the case in absence of these market imperfections.

For the purpose of our inquiry we are considering a dual economy, consisting of an advanced sector, and a subsistence one. Corresponding to the two sectors, there exist two occupations that individuals can choose between. The first occupation, advanced sector employment, requires the acquisition of a high level of skills, a typical example of which may be taken as a university degree. There are tuition fees so that for some individuals this option necessitates borrowing capital from the credit market. The alternative occupation does not require investments in human capital, and thus enables individuals to save on the fixed costs associated with entering the advanced sector, as well as the possible complications involved with borrowing. There is an information asymmetry in the credit market: borrowers have different levels of expected returns from education, and hence different probabilities of defaulting on their loans. While the distribution of returns is public information, banks cannot distinguish borrowers who are likely to default on the repayment of their loans from those who are not likely to.

We consider two concepts of poverty. Using an income concept, we set the poverty line at the payoff of the subsistence occupation. As defined by the classical economists, subsistence income is taken as the minimum income required in order to survive, or "subsist". The second concept we retain is one based on the human capital endowments of individuals. Schultz (1993), p.2 writes:

> "Where there is little human capital there is only hard manual work and poverty, except for those who have some income from property."

Individuals working in the subsistence occupation have low levels of education , and accordingly can be defined as poor on the basis of a human capital life-cycle theory of earnings. We shall be using the head-count measure in order to quantify the level of poverty. For instance, using the income concept, the level of poverty is taken to be the fraction of the population earning subsistence income.

With these definitions in mind, we state the two objectives of the chapter. Our first purpose is to examine the extent to which credit market imperfections can be a cause of poverty. In order to investigate this question, we contrast the amounts of poverty our model would account for in a regime of full information and one of asymmetric information. The second purpose of this work is to assess the potential for government intervention in the direction of poverty reduction. We show that depending on the demand for workers with higher education, the presence of information asymmetries may reduce, or have a neutral effect on the incidence of poverty. Also, we find that in general the social objectives of reducing poverty and maximizing total surplus have to be traded off against one another. We conclude our work by stressing the importance of examining the likely effects of efficiency enhancing policies on the incidence of poverty.

The model we examine below is closely related to the work of Mankiw (1986), and Bernanke and Gertler (1990). The three models have in common the adverse selection problem in the credit market, with respect to the default probabilities of borrowers. The major difference between our work and the previous two, is the emphasis placed on poverty reduction as opposed to the more frequently discussed efficiency objectives. The work of Starrett (1976) also attempts to establish a relationship between information problems and the occurrence of poverty. He constructs a model where "Economic agents somehow adopt the wrong model of the world, and their beliefs become self-reinforcing ." In such models poverty is endogenous to the system in the sense that

> "... an individual who is poor (or black , disadvantaged etc, ) may perceive his chances for success as small, and therefore not apply himself very much; he winds up a failure and appears to reinforce the view that his peer group is at a disadvantage even though the system is perfectly fair."

In Starrett's work poverty thus arises as a *rational expectations* equilibrium. Though uncertainty is common to our work and that of Starrett's, it is asymmetric information which drives our analysis, as opposed to the distorted beliefs leading agents to " adopt the wrong model of the world". Barham et al. (1992) study the relationship between education and poverty. However, their work centres on the theme of funding education through family channels exclusively, and rules out the possibility of borrowing through the credit market . As such, their theory is closer in essence to the work of Becker and Tomes (1986). Galor and Zeira (1993) also examine the allocation of credit in an economy with information asymmetries. Their purpose is to examine the long run evolution of the distribution of wealth and its impact upon long-run growth. They offer an analysis of a moving picture, close in spirit to the literature on growth theory. Other models similar to that of Galor and Zeira include Piketty (1992), Aghion and Bolton (1991), and Banerjee and Newman (1991).

Our chapter is structured as follows. In section 3.2 we present the framework we shall be using in order to explore the relationship between poverty and information problems. In section 3.3 we examine the equilibrium of the economy under a regime of full information. In section 3.4 we model the economy in presence of information asymmetries. In section 3.5 we contrast the amounts of poverty that result from both information regimes. In section 3.6 we examine the scope of government intervention for poverty reduction purposes. We conclude the chapter in section 3.7.

## 3.2 A Framework

We consider an economy with two sectors: a subsistence sector and an advanced one. The advanced sector is characterized by profit maximization, so that labour is hired up to the point where its marginal product equals the market wage. Meier(1984) p.162, defines the subsistence sector as being:

"... that part of the economy which does not use reproducible capital and does not hire labor for profit-the indigenous traditional sector or the self-employment sector."

In order to be able to enter the advanced sector, agents have to make an investment in human capital during the first period of their life, which would allow them to earn advanced sector wages w in the next period. This investment is associated with a risk factor, in the sense that each individual faces a specific probability 1-p of not being successful in his training. As an alternative to investing in human capital, an agent may choose to enter directly the subsistence sector in the beginning of the first period, and earn a wage c in each of the two periods. These occupational choices and their respective remunerations are drawn below.

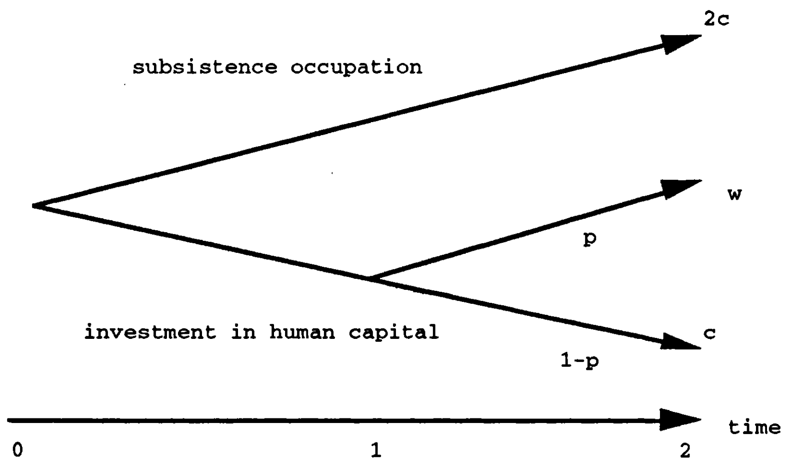


FIGURE 3.1: Remunerations from the alternative occupations

There exists a credit market for the purpose of financing education. The loan will typically cover tuition fees. The loan can also be viewed as comprising a living allowance b<c, in addition to the tuition expenses. If education contains consumption good aspects, than b can also be interpreted as the the non-monetary benefits of education. Following either of these interpretations, we can normalize the size of the loan at unity. The credit market is operating under a competitive regime , so that banks are willing to lend to potential students as the long as the expected rate of return on the loans equals the risk-free interest rate. The risk-free interest rate is exogenous and is set equal to zero, so that the supply of credit becomes driven by a zero-profit condition. That is, banks are willing to offer loans to students at an interest rate which allows them to *break even*.

Individuals in the economy differ in two respects: whether or not they have to use the credit market in order to fund their education, and their expected returns from the high skills occupation. For the purpose of our model , we distinguish three categories of individuals: F, NH, and NL. The proportions of individuals in each of the three categories is given in the table below.

| individual | F | NH | NL |
|---|---|---|---|
| proportion | $\phi$ | $(1-\phi)\lambda$ | $(1-\phi)(1-\lambda)$ |

Table 3.1: Categories of agents, and their respective proportions

Funded individuals (F) are those who do not require to use the credit market. Their funds can come from subsidized credit, bequests, or family funds available for the childrens' education[1]. To simplify the model, we assume that banks do not hand out cash to students, but rather pay the tuition and other fees directly to the school/college etc. This is a reasonable assumption since it can only reduce the riskiness of lending from the banks' point of view. We will also restrict ourselves to the situation whereby F-type agents are the ones who have no incentive whatsoever to use the credit market. This is the main feature which distinguishes them from the other two groups of individuals.

Individuals who require the use of the credit market are non-funded and come in two types: NH and NL. NH has a high success probability $p_H$ in his human capital investment, whereas NL has a lower success probability $p_L$. Since in the adverse state, workers only earn subsistence income from their study, they are obliged to default on their loan. Thus, the success probability of borrowers matters to the bank since it is also the probability of returning the

(1) Regarding the question of family funding of education, see Becker and Tomes (1986), and Barham et al. (1992).

loan. When the borrower's type is not observable to the bank, an adverse selection problem occurs in the credit market.

The demand for skilled workers is a function LD(w,q), where q is a shift parameter. The market wage is thus an endogenous variable. The borrowing price r, of the tuition fees is also an endogenous variable. It is determined by the supply and demand for credit. The model thus aims at explaining the market wage w , the terms of borrowing r, and the fraction of individuals who choose to acquire education. From the model we can deduce the share of individuals who are in poverty in terms of life-cycle income or human capital.

### 3.3 The Full Information Economy

Since we intend to examine the consequences of asymmetric information on the incidence of poverty, it is important to know what kind of equilibrium the economy achieves under a situation of full information. We therefore begin our discussion in the hypothetical case in which banks can observe the default probabilities of borrowers. Throughout our work we assume that individuals are risk neutral, so that utility levels can be expressed as linear functions of the returns arising from the various occupations.

Consider first NH and NL. They are those who have to use the credit market in order to finance their studies. NL has the low success and repayment probability $p_L$. His expected return from education is $b+p_L(w-r_L)+(1-p_L)c$, where $r_L$ is the repayment requested by the bank in exchange for covering NL's tuition fees. If he chooses the subsistence occupation, he can work as an unskilled worker and earn the cumulated wage 2c. He thus chooses to study if $b+p_L(w-r_L)+(1-p_L)c \geq 2c$. Since the credit market is operating under a zero-profit condition, the bank will set $r_L=1/p_L$. The reservation wage of NL (the minimum wage that induces NL to acquire education), is therefore:

$$w^*_{NL} = c + (c - b + 1)/p_L \qquad (3.1)$$

A non-funded student of type H, has a repayment probability $p_H > p_L$. Since the bank can identify an NH-type borrower, it sets $r_H = 1/p_H$. It follows that NH has a reservation wage $w^*_{NH}$ is such that:

$$w^*_{NH} = c + (c - b + 1)/p_H \qquad (3.2)$$

Let us now turn to the case of funded individuals. F is able to secure funding through a family transfer which he promises to return at a later date. The family acts as a bank: it requests $1/p_F$ if F is successful, and zero otherwise. F is therefore indifferent between using his family funds and borrowing from the credit market at a rate $r_F = 1/p_F$. Assuming families are also risk-neutral, the transfer then induces a neutral effect on their welfare. In behavioural models of the family, it is assumed that parents make direct investments in the child's human capital (see for example Becker and Tomes (1986)). In such models, a tradeoff arises between the family's consumption and the child's *quality*. In the above sense, our assumption that family transfers are neutral, presents a departure from the economic models of the family(2).

Since F does not borrow from the credit market, we can choose any value for the probability $p_F$ without changing the essence of the model. We let $p_F = p_H$. The reservation wage of F is therefore equal to that of NH:

$$w^*_F = w^*_{NH} = c + (c - b + 1)/p_H \qquad (3.3)$$

From equations 3.1-3 we can derive the labour supply schedule of the full information economy:

(2) We can also view F as having wealth endowment $m \geq 1$. Because of the risk-neutrality assumption ( implicit in the choice of the linear utility function), the exact value of m (provided $m \geq 1$) will generally not matter to the reservation wage of F. Alternatively, we can view F has having obtained a state loan at the risk-free interest rate.

$$LS(w)=\begin{cases} 0 & w < c+(c-b+1)/p_H \\ \in [0, E_H] & w = c+(c-b+1)/p_H \\ E_H & c+(c-b+1)/p_H < w < c+(c-b+1)/p_L \\ \in [E_H, E_H+E_L] & w = c+(c-b+1)/p_L \\ E_H+E_L & w > c+(c-b+1)/p_L \end{cases} \tag{3.4}$$

where

$$E_H = p_H\ [\phi + \lambda(1-\phi)] \tag{3.5}$$

and

$$E_L = p_L\ [(1-\lambda)(1-\phi)] \tag{3.6}$$

$E_H$ denotes the maximum endowment of educated labour of the F and NH populations. There are $\phi$ individuals of type F, and $(1-\phi)\lambda$ of type NH, each having a success probability $p_H$. By the law of large numbers, the maximum endowment of educated labour arising from these two groups will be $p_H\ [\phi + \lambda(1-\phi)]$, which we have defined in (3.5) above as the quantity $E_H$. Likewise, $E_L$ can be interpreted as the maximum endowment of NL-type skilled work-force.

The labour supply schedule consists of two steps and is drawn in figure 3.2 below. Let the demand for skilled labour be a function LD(w,q). The labour market equilibrium occurs at the intersection of the supply and demand schedules. In the diagram we depict a situation where the labour market clears at the reservation wage of high return individuals, i.e. at $w=(c-b+1)/p_H +c$. When such a situation arises, $\gamma/p_H$ high returns individuals acquire education. Of them, $\gamma$ individuals are successful and earn the wage $(c-b+1)/p_H +c$, (which is higher than the income poverty line 2c). Thus, while in ex-ante terms all the high returns individuals are indifferent between either occupations, ex-post only $\gamma$ workers are able to escape poverty. The remaining $1-\gamma$ workers in the economy join the subsistence occupation and are therefore poor in terms of either life-cycle income or human capital levels.

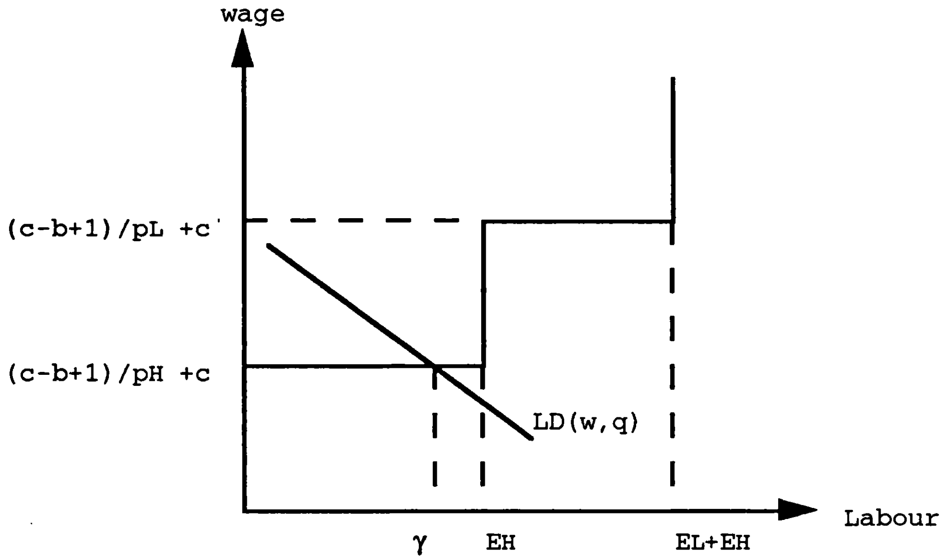


FIGURE 3.2: Labour market equilibrium in the Full Information Economy

Under a regime of full information, the market equilibrium has the important property of being Pareto efficient. By this we mean that the price mechanism decentralizes the economy in a way that social surplus is maximized. If fewer than $\gamma$ persons (selected from F and NH types ) invest in human capital, there is under-investment in education at the social level. Conversely, if a fraction greater than $\gamma$ invest in education, there will be some socially inefficient investments that are undertaken. Finally, note that at the wage $(c-b+1)/p_H+c$, there are no NL-type individuals investing in education. If a social planner were to place a low return student in education, the expected surplus from this decision would be $p_L(c-b+1)/p_H+c+b$. If NL were directed to the low skills occupation, social surplus would consist of one unit of money invested at the risk-free interest rate (zero) , plus the cumulated wage 2c of a subsistence sector worker. Social surplus under the latter choice is higher than under the former, by an amount $(c+1-b)(1-p_L/p_H)$.

### 3.4 Equilibrium Under Asymmetric Information

The aim of this chapter is to examine to what extent asymmetric information problems may contribute to the incidence of poverty. This section extends the discussion to deal with the adverse selection problem that arises in the credit market when the repayment probabilities of borrowers are not observable to banks.

Our starting point is to note the difference in the nature of equilibrium when the economy is operating in a regime of full information and asymmetric information. As pointed out by Akerlof(1970), markets operating under asymmetric information may have no equilibria. In the context of our model, the "lemons principle" operates on the demand side of the market since the quality of borrowers is uncertain to the suppliers of funds.

If there exists an equilibrium, it may come under two types: the pooling, and the separating types. A separating equilibrium can obtain if there exists an activity in which NH may invest in order to signal his identity, which would be too costly for NL to undertake. Such an situation would be feasible if the returns to higher education were correlated with an observed characteristic of borrowers. Typically, wealth is often advocated to be a correlate of ability. Separating equilibria could for instance be envisaged if we assumed that NH could signal his type by requesting to borrow only a fraction of the loan, $k<1$, while NL could not costlessly mimic NH's behaviour because of his smaller wealth endowment. Nonetheless, since the assumption regarding the correlation of ability and wealth is highly controversial, and difficult to verify, we will do away with the possibility of signalling activities, and shall restrict ourselves to the investigation of pooling equilibria.

Let $\pi(r,w)$ denote the average repayment probability of borrowers when the banks set the repayment value of the loan to r and the market wage is w. $\pi$ is the expected value of p for those persons who finance their education through the credit market:

$$\pi(r,w) = E[\, p \mid b+p(w-r)+(1-p)c \geq 2c \,] \qquad (4.1)$$

Equation (4.1) defines the demand for credit. The supply of credit is governed by the zero-profit condition of competitive markets:

$$\pi r - 1 = 0 \qquad (4.2)$$

where the size of the loan is set equal to unity. If the market wage is w and the cost of the loan is r, NH has an expected return from education $b+p_H(w-r)+(1-p_H)c$. He therefore borrows and invests in his human capital if $r \leq w+(b-c)/p_H-c$. Likewise, NL borrows if $r \leq w+(b-c)/p_L-c$. Therefore,

$$\pi(r,w)=\begin{cases} p_H & \text{if } w-c-(c-b)/p_L < r \leq w-c-(c-b)/p_H \\ \\ \lambda p_H + (1-\lambda)p_L & \text{if } \quad r \leq w+(b-c)/p_L-c \end{cases} \qquad (4.3)$$

where $\lambda p_H + (1-\lambda)p_L$ is the average repayment probability of the non-funded population. Below we define this quantity as $E(p)$. When $r>w-c-(c-b)/p_H$, no one is interested in borrowing and the credit market shuts down. When the return to investment is not high enough to induce any borrowing , there is a situation of "financial collapse" (Mankiw(1986)), the limiting case of what Bernanke and Gertler(1990) call "financial fragility". Equating (4.2) and (4.3), we obtain the following two equilibria of the credit market:

$$r=1/p_H \quad \text{for } c+(c-b+1)/p_H \leq w< c+1/p_H+(c-b)/p_L \qquad (4.4a)$$

$$r=1/[\lambda p_H + (1-\lambda)p_L] =1/E(p) \quad \text{for}$$
$$w\geq c+(c-b)/p_L + 1/E(p) \qquad (4.4b)$$

Let $\boldsymbol{W}$ denote the wage interval:

$$\left[c+1/p_H+(c-b)/p_L \, ; c+(c-b)/p_L+1/E(p)\right] \qquad (4.5)$$

We wish to inquire if for any wage $w \in \boldsymbol{W}$ the credit market admits equilibria with various proportions of high and low returns borrowers. If all the NH-type individuals, and a proportion $0 \leq \alpha \leq 1$ of the NL population borrow, then the average repayment probability must be

$$\pi = \left[ \lambda p_H + p_L(1-\lambda)\alpha \right] / \left(\lambda + (1-\lambda)\alpha\right) \qquad (4.6)$$

For low returns individuals to borrow when the lending rate is r, the market wage w must be high enough such that $b+p_L(w-r)+(1-p_L)c \geq 2c$. Since, by the zero-profit condition, $r=1/\pi$ , the reservation wage of an NL-type borrower is then:

$$w^*_{NL} = c + (c-b)/p_L + \left(\lambda + (1-\lambda)\alpha\right) / \left[\lambda p_H + p_L(1-\lambda)\alpha\right] \qquad (4.7)$$

when $\alpha$->0, there are only high returns borrowers, so that $w^*_{NL}$ approaches the lower bound of the wage interval ***W***. Likewise, when $\alpha$->1, all the low returns individuals borrow, and $w^*_{NL}$ approaches the upper bound of the wage interval ***W***. The intuition behind the result is the following: for the credit market to accommodate greater numbers of low return borrowers, a rise in the interest rate is required in order to compensate lenders for the increase in default rates. When the lending rate rises, so must the reservation wages of borrowers increase.

Solving for $\alpha$ in (4.7) we have

$$\alpha(w) = \frac{\lambda}{1-\lambda} \left[ \frac{p_H\{w - c - (c - b)/p_L\} - 1}{1 - p_L\{w - c - (c - b)/p_L\}} \right] \qquad (4.8)$$

so that for $w \in \boldsymbol{W}$, labour supply of non-funded low returns individuals equals $\alpha(w)E_L$ .

We now proceed to construct the labour supply schedule of the economy within the context of the asymmetric information regime. We first note that F does not have any incentive to borrow: if $r=1/p_H$ he is indifferent between borrowing and using his own funds, when $r>1/p_H$ he never borrows. His reservation wage therefore remains unchanged at $c+(c-b+1)/p_H$. The reservation wage of NH also remains unchanged: when $r=1/p_H$, the lowest wage that will induce him to invest in education is also $c+(c-b+1)/p_H$. On the other hand, as a result of the inability of banks to distinguish between borrowers, NL can borrow at the interest rate $r=1/\pi$, as opposed to $1/p_L$ within the full information scenario. The lowest wage that may induce an NL type person to borrow is given by (4.7), which for all values $0 \leq \alpha \leq 1$, is lower than the corresponding value under the full information assumption.

It is worth asking how the reduction in NL's reservation wage arises. When the demand for skilled labour is relatively high, the credit market can allocate funds to cover agents with varying degrees of returns to education. The credit market then implicitly redistributes income from individuals with above average returns, to those with below average returns to education. That is, NL's reservation wage drops, because through the market allocation of credit NH implicitly subsidizes the former's loan.

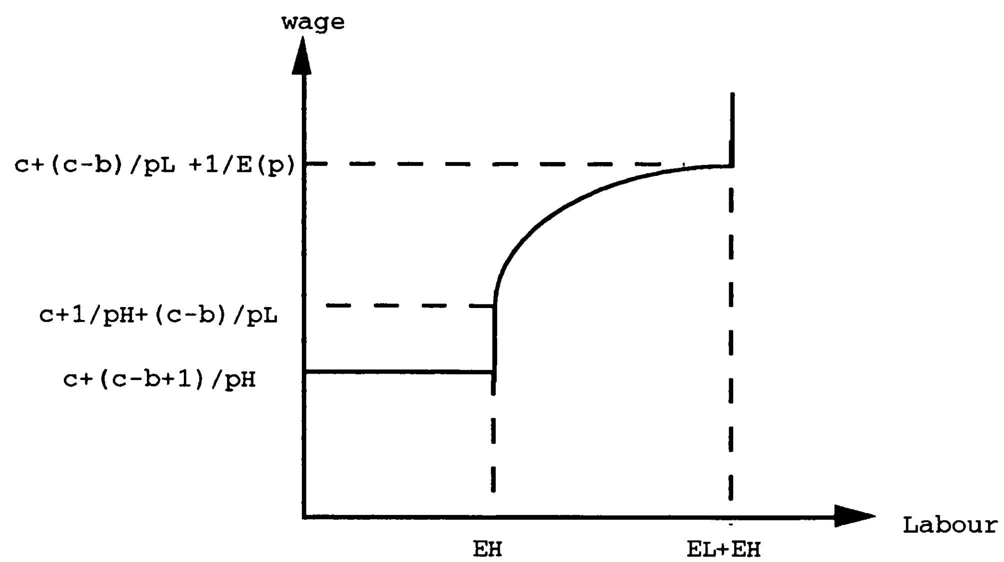


FIGURE 3.3:Labour Supply under Asymmetric Information

The labour supply schedule can therefore be constructed as follows:

$$LS(w)=\begin{cases} 0 & w < c+(c-b+1)/p_H \\ \in [0, E_H] & w = c+(c-b+1)/p_H \\ E_H & c+(c-b+1)/p_H < w < c+1/pH+(c-b)/p_L \\ E_H+\alpha(w)E_L & w \in \boldsymbol{W}, \\ E_H + E_L & w \geq c + (c-b)/p_L+1/E(p) \end{cases} \qquad (4.9)$$

where $\boldsymbol{W}$ is the wage interval defined in equation (4.5) above. Differentiating $\alpha(w)$ in (4.8), we note that its first and second derivatives are positive for all wages $w \in \boldsymbol{W}$. The labour supply schedule is sketched in figure 3.3 above.

As usual, the labour market clears at intersection of the supply and demand schedules. The consequences of asymmetric information on the incidence of poverty are spelled out in the section below.

### 3.5 Consequences for Wages and Poverty

One of the aims of this work is to characterize the effects of credit market imperfections on the incidence of poverty. As a starting point in our inquiry, we superpose the full information and the asymmetric information labour supply schedules of figures 3.3 and 3.4.

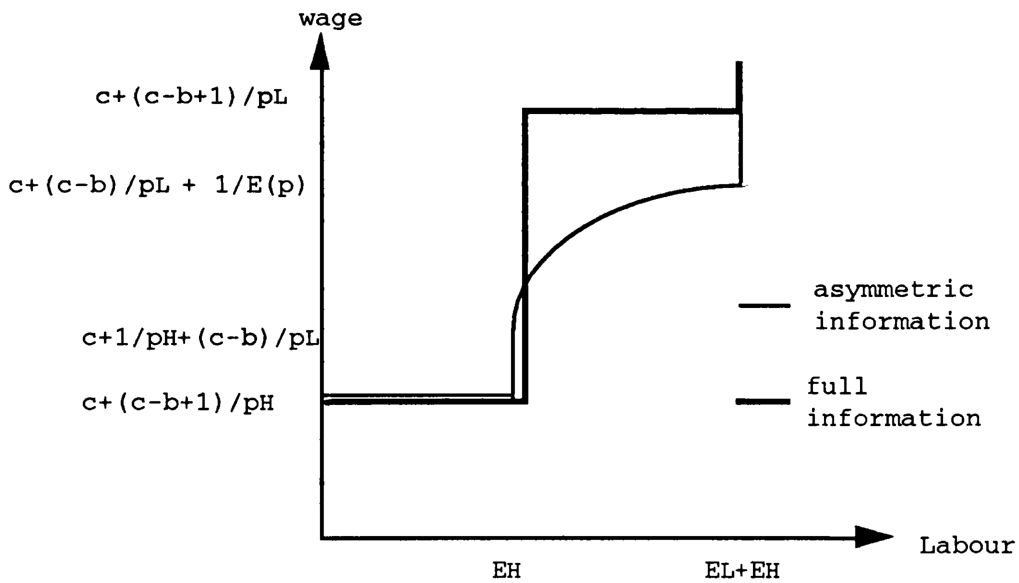


Figure 3.4: Labour supply under both information regimes

As we shall see shortly, the consequences of asymmetric information on the incidence of poverty, will depend on the level of demand for skilled labour. Implicit to our analysis is the assumption that production in the advanced sector is characterized by decreasing returns, and that the price of the output is exogenously determined (as would be the case for a small open economy). We denote the labour demand schedule by LD(w,q), so that shifts in the demand for labour can be indexed by changes in the parameter q. We will consider four cases below, depending on the level of the shift parameter q.

**case 1**: $q_o < q \leq q_1$

We note first that for there to be an advanced sector in the economy, the demand for labour must at least attain $LD(w,q_o)$ so that the equilibrium wage attains the minimum value $c+(c-b+1)/p_H$.

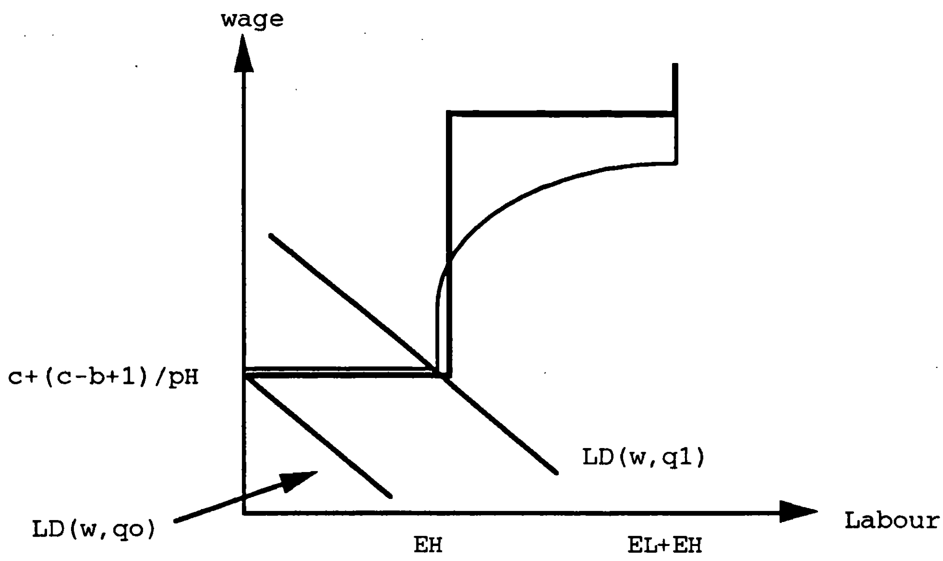


FIGURE 3.5 : The Low Wage Economy

Below $LD(w,q_o)$ the modern sector would not be viable and the economy would devote all its productive resources to subsistence activities. When $q_o<q\leq q_1$, the expected payoff to advanced sector workers is equal to 2c. In an ex-ante sense, we are in the context of an income-poor economy. Ex-post though, those individuals who are successful in their

training earn the wage $c+(c-b+1)/p_H$ and therefore cross the income poverty line.

In the low wage economy, the demand for skilled labour is not high enough to induce NL to study. Thus, NH borrows on the same terms as he would in a regime of full information. Both labour supply curves overlap, so that the asymmetric information regime leaves wages and employment in the high skills sector unchanged. This result can be explained by noting that when wages are too low to attract NL, there is no information asymmetry as such between borrowers and lenders. The market equilibrium of the low wage economy is identical under both information regimes, and so is the incidence of poverty.

**case 2**: $q_1 \leq q \leq q_2$.

When the demand for labour rises to the intermediary region, the net payoff to higher education rises above 2c. On the basis of either income or human capital concepts, the poverty count is $1-E_H$ since all successful high returns workers are absorbed by the advanced sector.

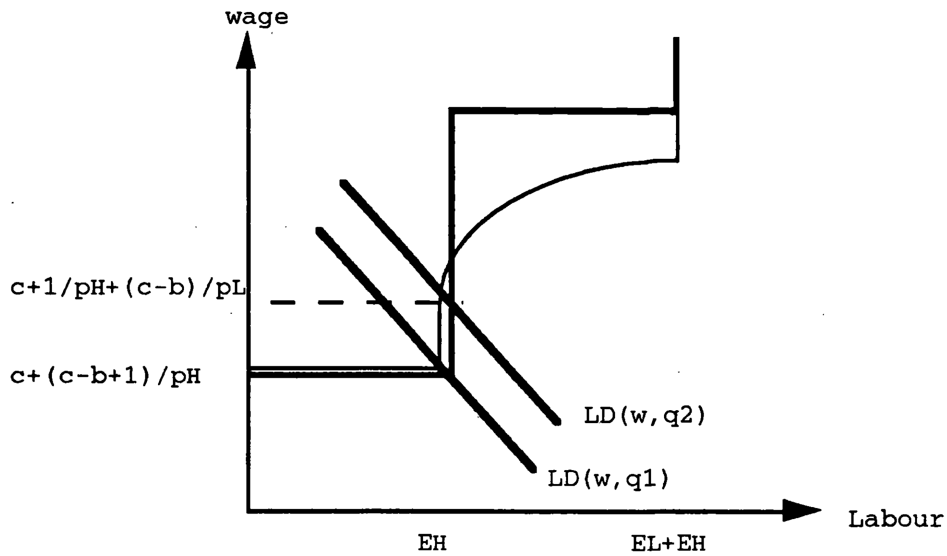


FIGURE 3.6: The intermediary economy

Since in the intermediary economy wages remain too low to attract low return borrowers, both labour supply curves overlap. Wages and poverty therefore remain identical under both regimes.

**case 3** $q_2 < q \leq q_3$ .

By a high wage economy, we mean that wages are high enough to attract low returns workers to the advanced sector.

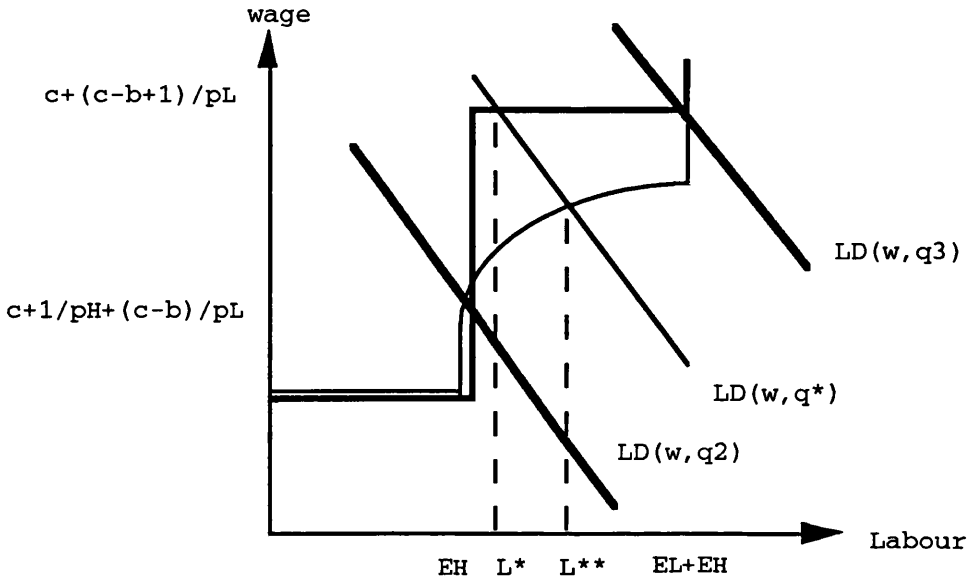


FIGURE 3.7: The high wage economy

As was explained in the previous section, the reduction in NL's reservation wage is achieved via an implicit subsidy of his loan, originating from NH.Because the asymmetric information labour supply schedule lies below the full information one (in the region $q_2 < q \leq q_3$), wages will be lower and employment higher as a result of the credit market imperfections. When the demand for labour is LD(w,q*)[3], the economy operating in a regime of full information would have a proportion 1-L* of its work-force in poverty. Within the context of the adverse selection regime, the incidence of poverty is reduced to 1-L**. It is important to note that from an efficiency point of view social surplus is maximised when L* individuals invest in

(3) Note that LD(w,q*) can intersect the two labour supply curves in four different ways. It can cut the full information schedule in its horizontal segment or its vertical segment. Likewise, LD(w,q*) can cut the asymmetric information labour supply in its curved segment and its vertical segment.

education. The reduction of poverty obtained in the asymmetric information regime therefore arises because of over-investment in education [4]. An NL-type person gains, whereas F and NL see education less profitable in the adverse selection regime than in the full information set-up of the high wage economy.

**Case 4**: $q > q_3$

When the labour demand schedule is situated above LD(w,q3), all the labour force invests in education. Such an economy has reached its full productive capacity under both information regimes. Further increases in the demand for skilled labour result in higher wages but no further supply of workers. In this sense the economy can be said to have reached full-employment.

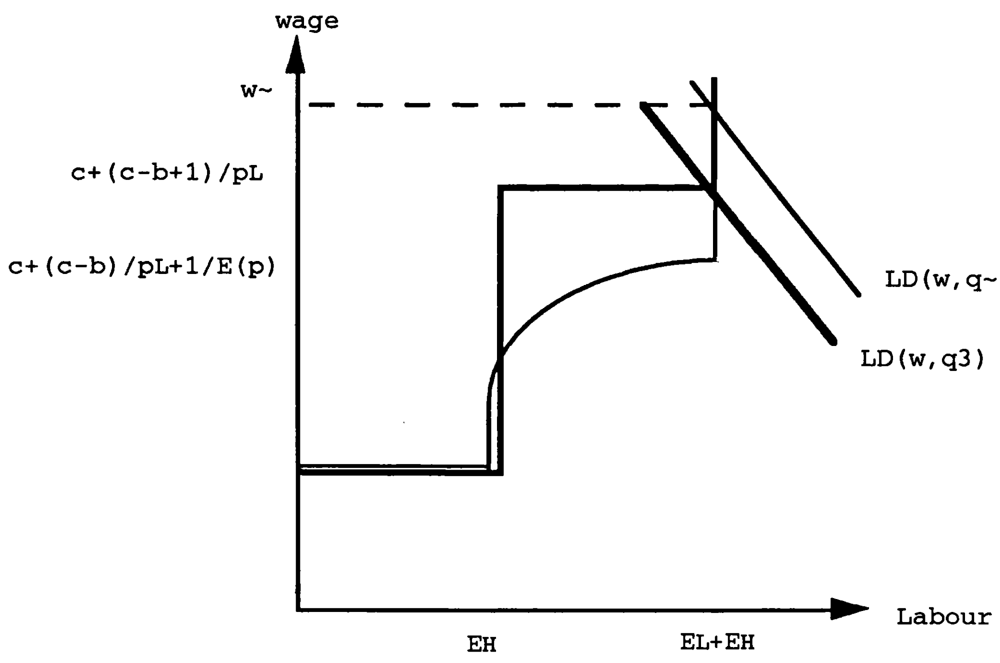


Figure 3.8 : The Full-Employment Economy

The resulting market wage exceeds the reservation wage of NL within the context of a full information regime, and therefore both labour supply schedules overlap. As in the case of the low wage and intermediary economies, the

---

(4) This is a standard result in screening models of the labour market. See for instance Spence (1973).

case of the low wage and intermediary economies, the existence of asymmetric information in the credit market has a neutral effect on wages and poverty . This is illustrated in figure 3.8. When the labour demand is LD(w,q~), the resulting wage is w~ under both regimes.

The conclusion that emerges from this section is that the adverse selection problem in the credit market will generally not result in higher levels of poverty. The low wage and intermediary economies are characterized by sufficiently low levels of demand for skilled labour, so that wages and poverty are identical under both information regimes. High wage economies on the other hand, achieve lower levels of poverty in presence of the adverse selection regime. In the limiting case of the full-employment economy, both information regimes entail the same outcome with respect to wages and poverty.

**3.6 Implications for Poverty Reduction Policies**

Since education is an important channel by means of which individuals can invest in their human capital, the existence of asymmetric information problems in the economy gives rise to a substantial role for government policy on education and poverty. The purpose of this section will be to examine with the help of our model , the scope for government intervention in the labour market for the purpose of reducing poverty. Other equally important social objectives on education relate to efficiency and equality of opportunity matters.

Our discussion will be mainly centred around the question of poverty reduction. Efficiency questions have been dealt with in Mankiw (1986), Bernanke and Gertler (1990), and in Hoff (1991). The relationship between poverty, education, and equality of opportunity was dealt with, in Bowles (1973).

If the government wishes to pursue poverty reduction policies it can intervene in either of the credit or labour markets. Credit subsidies could lower the reservation wages of workers. Equivalently, wage subsidies could result in higher levels of employment in the advanced sector. Here we limit ourselves to intervention in the labour market.

One way of reducing poverty would consist in using profits in order to subsidize wages. Since we have assumed a decreasing returns to scale technology, it follows that the average product of labour, G(L,q)/L, exceeds its marginal product at any level of employment

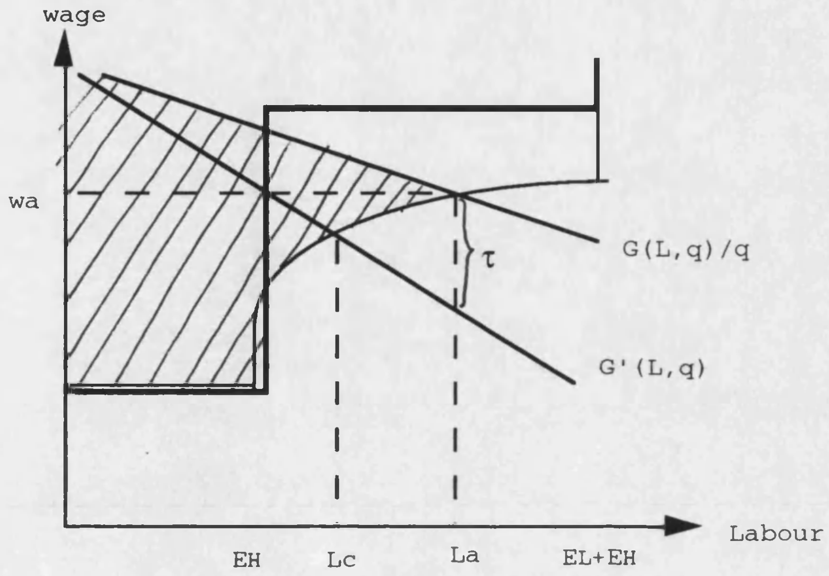


Figure 3.9: Poverty Minimum

The shaded region in figure 3.9 represents the feasible set of wages and employment the advance sector can reach given its labour force and its production technology G(L,q). For a given level of employment, wages cannot exceed the average product of labour, since then profits would be negative. Conversely, to induce participation of workers, wages must not fall short of the reservation wage corresponding to the given level of employment. Thus, any pair (L,w) must be chosen inside the shaded area.

A social planner concerned with poverty minimisation would therefore choose to set employment at its highest possible level. The point $(L_a,w_a)$ where the labour supply and average product of labour curves intersect is the maximum employment level attainable in the advanced sector, and thus the minimum *feasible* poverty level. The unregulated market equilibrium on the other hand has $L_c$ workers in the advanced sector. The poverty reducing policy therefore consists in subsidizing wages by whatever profits may be available. At $L_a$, the marginal product of labour is lower than $w_a$, so that a subsidy $\tau$ (per worker) is required to make up for the difference. Such a policy would allow a further reduction of poverty by an amount $L_a - L_c$.

It is important to note that only under a constant-returns-to-scale assumption will the market equilibrium of the (asymmetric information regime) automatically attain the poverty minimum. This is because when there are constant returns to scale, marginal and average products coincide and profits are null. In presence of decreasing returns, government intervention in the form of profit taxes and wage subsidies, will allow further reductions in poverty over the unregulated market equilibrium.

As can be seen from figure 3.9, poverty reduction has to be traded off against efficiency losses. The efficient level of employment obtains at the intersection of the marginal product of labour and the labour supply of the full information regime, i.e. at EH. Poverty reduction is obtained through the combined effect of wage subsidies and profit taxes, which induces over-investment in education. Efficiency considerations therefore call for a contraction of employment in the advanced sector. Concerns about poverty on the other hand, call for an expansion of employment.

### 3.7 Conclusions

It is well understood by now that information asymmetries in the credit market may distort the allocation of investment from the social optimum. This work has examined the consequences of asymmetric information on poverty. We have shown that credit market imperfections may have a neutral effect on poverty, but they may also reduce it. The latter possibility was shown to arise in *high wage economies*. There, through the market allocation of credit, individuals with above average returns to education subsidize those with below average returns.

A conclusion that emerges from the work, is the importance of assessing the likely effects of efficiency enhancing policies on the incidence of poverty. One consequence of information asymmetries in the credit market, is the resulting over-investment in education that may arise in the market equilibrium. Policies aimed at raising social surplus will then reduce investment in education, and hence will result in poverty increases. It is thus important that poverty reduction does not escape unnoticed as a goal, especially when efficiency gains have to be traded off against poverty increases.

CHAPTER 4

# A Justice Criterion Pertaining to Intergenerational Mobility Processes

## 4.1 Introduction

One important conclusion that emerges from the model of the previous chapter, is the importance of family background in the determination of income. In chapter 3 we saw that non-funded individuals with high returns on their human capital investments (NH) were at disadvantaged with respect to their funded counterparts (FH), since high return borrowers implicitly had to subsidize the loans of low return borrowers. Given these differing initial circumstances of individuals, what distribution of income today constitutes a just state of affairs?

The question we are dealing with in this chapter can more generally be formulated within the context of the normative assessment of intergenerational mobility processes. Given a joint distribution of income f(x,y) for parents and children pairs, we wish to decide whether the given outcome characterizes a just state of affairs. We divide the population of children into various groups, according to their socio-economic backgrounds, and suggest to qualify the process f(x,y) as *un*just or *un*fair[1], if any single one of these groups enjoys a superior distribution of welfare than any other one.

We begin our discussion with the following question: if say 10% of the population today are poor, would it be irrelevant if these individuals all originated from low income families, or would it be a preferred state of affairs if these individuals had purely random socio-economic backgrounds? We take the view that *who* low income earners

---

[1] Here we take justice and fairness as synonyms. In social choice theory, fairness is associated with the concepts of envy and equitability. See Hammond (1987) for a discussion.

are today matters, not just how many of them there are. Presumably if asked to choose, most of us would perceive the second state of affairs, where the poor are of mixed origins, as being preferable to the case where they all originate from poor families.

If this is indeed the case, we need to further inquire as to why the socio-economic origins of individuals should be of relevance in the normative assessment of intergenerational mobility processes. Atkinson et al. (1983) pp. 14-15, argue that when individuals live for several periods, a high degree of intergenerational mobility may be viewed as instrumental to securing a higher level of equality of life-time incomes. For example, assume that individuals live for two periods, first as children, then as parents. A high degree of intergenerational mobility would ensure that those who had low living standards in their childhood, would thus not be further penalized during their parenthood, and that the benefits of being born wealthy are not further expanded over the life cycle.

Broadly speaking, we can distinguish two approaches with regards to the conception of economic justice. In the first approach the emphasis is laid on studying the *processes* by which incomes are determined and goods are allocated to the different members of society. In the second approach the assessment of justice is made on the basis of the *outcomes*, or *results*, of the market process.

In the context of the subject of the present work, following the first approach would lead us to study the process by which the incomes of parents are mapped onto those of their children. Typically, equality of opportunity arguments arise from this stance. If say, discrimination in access to education and employment are at the cause of the observed pattern of intergenerational mobility, then we may argue that justice is violated. In the first approach it is therefore actions of individuals as opposed to outcomes which form the basis of justice assessment. In the words of Hayek (1982), justice would proceed,

"...according to rules guiding the actions of individual participants whose aims, skills, and knowledge are different, with the consequence that the outcome will be unpredictable and that there will regularly be winners and losers. And while, as in a game, we are right in insisting that it be fair and that nobody cheat, it would be nonsensical to demand that the results for the different players be just."

The second approach stands in contrast to the former, in the sense that it is solely the results of actions which guide our judgements. An often invoked principle related to the outcomes approach, is that of consequentialism. In that context, Sen (1988) writes:

"...all choice variables, such as actions, rules, institutions, etc, must be judged in terms of the goodness of their respective consequences."

A well known justice concept which follows from the outcomes school, is Rawls'(1971) *difference*, or *maximin* principle. According to the difference principle, incomes and other fundamental goods, are to be distributed equally, and inequalities should only be tolerated in the case that they raise the welfare of the least well-off members of society.

In proposing a justice criterion related to intergenerational mobility processes, we shall be following the outcomes approach. Our justice criterion is a judgement on the income distributions of children originating from different socio-economic groups. Let us temporarily go back to the justice assessment of single income distributions, as opposed to the joint distribution of incomes of parents and children. In the context of a univariate distribution, it is common in the outcomes approach to associate justice with a scenario of equal incomes. The maximisation of a sum of increasing and concave identical utility functions, under the assumption that national income is constant, has for solution the equal sharing of the cake(2).

Let us divide the population of children into two socio-economic groups: the first group is that of the disadvantaged ones, and the second one consists of the privileged. The classification is based on the socio-economic status of parents, which we also refer to as the family background of children. Each of these groups is associated with an income distribution $f(y|X_i)$, where $X_i$ indexes the socio-economic origin of the child. Over which dimension should the principle of equality be defined in the analysis of groups of individuals as opposed to single units? In moving from a unique set of individuals to the analysis of groups of individuals, one possibility would be to replace the principle of equality of incomes by a principle of *equality of distributions.*

Equality of distributions for the two groups of children is statistically equivalent to a concept of independence from the socio-economic origin. To see that this is the case, note that the latter condition implies that $f(y|X_1)=f(y|X_2)=f^*$. That is, when distributions are equal, children raised in privileged and disadvantaged environments do equally well *on average.*

The concept of equality of distributions may be viewed as a rather strong requirement, and perhaps requires formulation in a more general context. One possible extension would be to replace the condition of equality of distributions by one of equality of aggregate welfare levels. Let $\boldsymbol{U}$ define the class of increasing and concave utility functions. Also assume that social welfare for group i is measured by the sum $\int u(y)f(y|X_i)dy$, where u is a member of $\boldsymbol{U}$. Whereas equality of distributions requires that $f(y|X_1)=f(y|X_2)$, equality of welfare obtains when $\int u(y)f(y|X_1)dy = \int u(y)f(y|X_2)dy$. Under the condition of *equality of welfare*, the two income distributions need not be identical. The income distribution for the children of

(2) Other generalizations of this result can be found for instance in Sen (1973).

disadvantaged backgrounds could for instance have a lower mean income than the corresponding distribution for the privileged children. But provided $f(y|X_1)$ is sufficiently more equally distributed than $f(y|X_2)$, the two distributions may achieve identical levels of aggregate welfare.

The principle of equality of welfare may appear once again fairly intractable in empirical applications. Agreement on whether justice obtains could only be achieved conditional on the full parametrization of the utility function u(). In this sense, a more flexible criterion could allow for different views on the form of the utility function. Let $u_a$ and $u_b$ be any two distinct members of the class $\boldsymbol{U}$ of utility functions, and consider the following three cases:

(I) For all $u \in \boldsymbol{U}$, $\int u(y)f(y|X_i)dy > \int u(y)f(y|X_j)dy$. That is, for any increasing and concave utility function, the children originating from group i are better off than those of originating from group j. Stated differently, there is a situation of *welfare dominance* in favour of the children from group i.

(II) $\int u_a(y)f(y|X_1)dy > \int u_a(y)f(y|X_2)dy$, while $\int u_b(y)f(y|X_1)dy < \int u_b(y)f(y|X_2)dy$. That is, depending on the choice of the utility function, it may be the case that children of disadvantaged origins are perceived as being better off than the privileged ones, and vice versa.

(III) Equality of distributions: $f(y|X_1)=f(y|X_2)=f^*$. All groups of individuals enjoy the same distribution of living standards regardless of their socio-economic origins.

Scenario (I) is a clear-cut case where the living standards of groups of individuals are determined by their socio-economic origins, and the welfare ranking of income distributions holds for any increasing and concave utility function. The justice criterion we propose in this work defines scenario (I) as a case of injustice. Under scenario (III) all groups enjoy identical living standards regardless of their socio-economic origins. Since all income distributions are equal, this conclusion must be valid for all $u \in \boldsymbol{U}$. We therefore take scenario (III) to be one where there is *absence of* injustice. Scenario (II) is the intermediary situation where we are not able to conclude which of the two distributions will entail a higher level of welfare without further knowledge on the form of the utility function u(). When such a situation arises, we are not able to qualify an intergenerational mobility process f(x,y) with respect to the presence or absence of injustice.

The outline of the chapter is the following. In the next section we present our proposed criterion for the justice assessment of intergenerational mobility processes. Our justice concept may be endorsed by certain theories of justice, but may also clash with others. We also take up this point in section 4.2. In section 4.3 we contrast our justice index with existing measures of intergenerational mobility. In section 4.4 we explain how the proposed index can be implemented with sample data. As shown by Shorrocks(1983), the ranking of income distributions can be carried out via the comparison of Generalized Lorenz curves for a wide class of social welfare functions. Since distribution-free methods exist for the estimation of Lorenz curves (see for instance Beach and Davidson(1983)), the conclusions derived on the basis of our justice index can be fairly general with regards to both the underlying social welfare function, and the parametric family of the income distribution. In section 4.5 we discuss some of the data limitations we face in the analysis of intergenerational mobility processes. We then proceed in section 4.6 to the empirical applications. Our final section concludes the chapter.

## 4.2 Justice and Welfare Dominance

We suggest that the potential for ranking the income distributions of the various socio-economic groups should constitute the basis of the justice assessment of intergenerational mobility processes. The socio-economic classification is based on the family background of individuals. For the purpose of ranking income distributions, we will restrict ourselves to welfare functions of the utilitarian type:

$$W( y_1,...,y_n ) = \Sigma u(y_i) \qquad (a)$$

so that the welfare function is anonymous, and the utility level of each individual depends solely on his income level y. We also assume that W exhibits a social aversion to inequality, so that mean-preserving progressive transfers are assumed to increase welfare. This amounts to a property of concavity of the individual utility functions[3] .

$$u''(y) \leq 0 \qquad (b)$$

In order to capture the desire for higher incomes, we will let W be a non-decreasing function of all incomes:

$$W( y_1,...,y_n ) \text{ is non-decreasing in } y_i \qquad (c)$$

The monotonicity of W(.) is one way of capturing social preference for efficiency. The properties (a)-(c) amount to a choice of a utility function which is increasing and concave, i.e. u() must be chosen from the set $\boldsymbol{U}$ of utility functions.

Let f(.) and F(.) respectively denote the density function, and the cumulative distribution function, of y. Also, define $Y=( y_1,...,y_n )'$ as the vector of incomes. The pth income quantile $\varepsilon_p$ is defined via the following equality.

$$F(\varepsilon_p) = p \qquad (2.1)$$

If say, p=0.5, then $\varepsilon_{0.5}$ is the median income in the population. Corresponding to a set of k horizontal ordinates $p_1<p_2<...<p_k$, we have a set of k income quantiles $\varepsilon_1<\varepsilon_2<...<\varepsilon_k$, more generally known as *order statistics*. The

---

[3] The condition can be weakened to a property of "S-concavity" of W , so that $W( B Y) \geq W(Y)$ for all bi-stochastic matrices B, where $Y=( y_1,...,y_n )'$.

Lorenz curve ordinate at p is defined as the fraction of mean income held by the bottom p percent of the population:

$$LC(p) = \frac{1}{\mu}\int_0^{\varepsilon_p} y\, f(y)\, dy \qquad (2.2)$$

The Generalized Lorenz curve, introduced by Shorrocks (1983), is defined as the Lorenz curve multiplied by mean income:

$$GLC(p) = \int_0^{\varepsilon_p} y\, f(y)\, dy \qquad (2.3)$$

Also, Let $\boldsymbol{W}$ denote the set of social welfare functions satisfying properties (a) to (c). Shorrocks (1983) proves the following theorem:

$$W(Y_1) \geq W(Y_2) \text{ for all } W \in \boldsymbol{W} \text{ iff } GLC_1(p) \geq GLC_2(p) \quad p=1,..,k$$

In other terms, the distribution $Y_1$ will be preferred to $Y_2$ for all $W \in \boldsymbol{W}$, if and only if $GLC_1$- the Generalized Lorenz curve of $Y_1$, lies nowhere below $GLC_2$ - the corresponding curve for $Y_2$. The comparison of Generalized Lorenz curves of various socio-economic groups will thus enable us to investigate the fairness of intergenerational mobility processes.

Define x as the income variable of the first generation which we refer to as *parents*. Also, let

$x \in X_1$ if the parents are disadvantaged

$x \in X_2$ otherwise

Aggregate welfare of children originating from class i is therefore:

$$W(y_1,...,y_n \mid X_i) = \int u(y) f(y \mid X_i) dy \qquad (2.4)$$

We suggest to qualify an intergenerational mobility process as unjust if either of the two income distributions $f(y|X_1)$ and $f(y|X_2)$ welfare dominates the other. Geometrically, the justice requirement can be stated in terms of Generalized Lorenz curves. Let $GLC_i$ denote the Generalized Lorenz curve for the children of socio-economic background $X_i$. Then an intergenerational mobility process $f(x,y)$ exhibits injustice if either of the two Generalized Lorenz curves lies nowhere below the other.

In figure 4.1 we consider two hypothetical cases. Under both scenarios the children from privileged origins enjoy a higher mean income than their counterparts from the other group. Under case (i) we have a scenario of injustice. That is, $W(y_1,...,y_n \mid X_2) > W(y_1,...,y_n \mid X_1)$ for all $W \in \boldsymbol{W}$ so that the distribution of income for children from privileged groups dominates that of the children from disadvantaged backgrounds. In other words, the advantages of being raised in privileged backgrounds tend to maintain themselves over the working life.

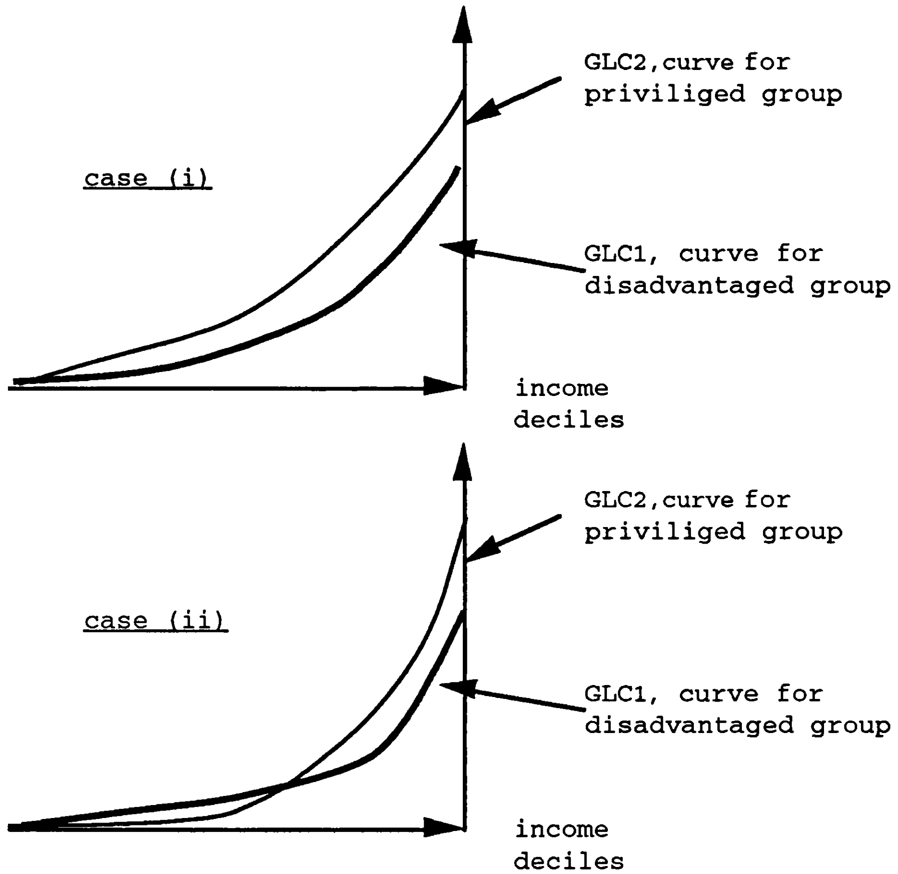


FIGURE 4.1: Justice and the Generalized Lorenz dominance criterion

Case (ii) on the other hand has the two Generalized Lorenz curves intersecting. That is, for some social welfare functions, the disadvantaged children would appear to be better off than the children from the other group. Under case (ii) one cannot unambiguously state that initial advantages are compounded over the life cycle, so that we cannot conclude whether there is a situation of injustice. We summarize these points in definition 2.1:

DEFINITION 2.1 : *Let there be two socio-economic classes:* $X_1$ *and* $X_2$. *Then*
*(i) An intergenerational mobility process* $f(x,y)$ *exhibits injustice if either* $f(y|X_1)$ *welfare dominates* $f(y|X_2)$ *or vice-versa if* $f(y|X_2)$ *welfare dominates* $f(y|X_1)$.
*(ii)* $f(x,y)$ *cannot be qualified with respect to the existence of injustice if* $f(y|X_1)$ *and* $f(y|X_2)$ *cannot be ranked in terms of welfare dominance.*
*(iii)* $f(x,y)$ *does not exhibit injustice if* $f(y|X_1)= f(y|X_2)$.

*A numerical example*

Let us illustrate the above statements with the help of a numerical example. Assume the joint distribution of income of parents and children is the following:

$$P=\begin{bmatrix} 1/12 & 0 & 5/12 \\ 0 & 3/12 & 3/12 \end{bmatrix}$$

where $p_{ij}$ denotes the proportion of (x,y) observations such that the parent belong to socio-economic class i, and the child belongs to socio-economic class j. Parents are classified into the two groups, i=1,2. Children on the other hand are classified into three income groups (also in increasing order). The marginal income distribution of fathers is obtained by summing up the columns over the appropriate line:

$$p_{i.} = \Sigma_j \, p_{ij}$$

where $p_{i.}$ denotes the probability that a father belongs to class i. Likewise,

$$p_{.j} = \sum_i p_{ij}$$

and from the laws of probability

$$\sum_j p_{i.} = \sum_i p_{.j} = \sum \sum p_{ij} = 1$$

The ijth element of $\Pi$, the transition matrix, is defined as the probability that the child is in income class j, given that the he originates from group i. From Bayes' theorem,

$$\pi_{ij} = p_{ij} / p_{i.}$$

so that each line of the transition matrix denotes a probability distribution. In our numerical example $\pi_{ij}=2p_{ij}$, i.e.,

$$\Pi = \begin{bmatrix} 1/6 & 0 & 5/6 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

To see that $\Pi$ cannot be qualified with respect to the existence of injustice, we note that if from society's point of view, the social utility of being in income class j is $u_j$, where say $u_1=1$, $u_2=3/2$, and $u_3=15/8$, then welfare is higher for the children originating from disadvantaged backgrounds. On the other hand, if say $u_1=1$, $u_2=3/2$, while $u_3=13/8$, the second line of the transition matrix dominates the first. In terms of Generalized Lorenz curves, we obtain a scenario similar to that sketched in case (ii) of figure 4.1. Now if we change the joint distribution of parent's and children's incomes in order to obtain the transition matrix $\Pi^*$

$$\Pi^* = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

Then for all $W \in \boldsymbol{W}$, the bottom line of $\Pi^*$ will always exhibit a higher level of aggregate welfare than the top line. Hence, as sketched in case (i) of figure 4,1, The GLC curve of children from privileged backgrounds lies nowhere below the corresponding curve for the children from disadvantaged backgrounds. In this sense, the intergenerational mobility process of $\Pi^*$ does not exhibit fairness. Finally, if both lines of the transition matrix are identical, we can reject the assumption that the underlying intergenerational mobility process exhibits injustice.

* * *

Note that definition 2.1 can be re-formulated within the context of $c > 2$ socio-economic groups. For instance we might be interested in looking at three groups, where say $X_2$ would index the children originating from middle classes while $X_3$ would pertain to those of the most privileged backgrounds. The three conditions would then be formulated as follows: (i) there is injustice if for any pair (i,j) the distributions $f(y|X_i)$ and $f(y|X_j)$ can be ranked in terms of welfare dominance, (ii) $f(x,y)$ cannot be qualified with respect to injustice if no single pair of distributions can be ranked in terms of welfare dominance and, (iii) $f(x,y)$ does not exhibit injustice if for all i $f(y|X_i)=f^*$, that is under a scenario of equality of all c conditional income distributions.

At this stage it is worth noting some of the limitations underlying our approach. In extending the analysis to incorporate c>2 income groups, a situation may arise whereby a conclusion is reached on the basis of c groups, and another one obtains using c+1 classes. Many of the mobility measures based on transition matrices also suffer from this problem (cf. section 4.3 below). Also, underlying the c income classes is a continuous variable, namely life-cycle income. In general, discretizing a continuous Markov Process may result in a poor approximation of the underlying true process (see Feller (1968), vol. I for a discussion). In this sense, adopting the *stochastic kernel* approach developed in Quah (1994) would be a useful extension in the normative assessment of intergenerational mobility processes.

Finally, we note that the definition of a generation is not a clear-cut time unit. A process may exhibit injustice on the basis of say, an 18 year time period separating the observation of the incomes of parents and children, but the role of family background in the determination of income may vanish when the length of a generation is taken as twenty five years. It would therefore be useful to examine the possibility of introducing a *period invariance* axiom (cf. Shorrocks (1978)) in the framework we have suggested.

Going back to the simple case of two socio-economic groups, when the justice condition is violated there are two scenarios of relevance. The first is one where $W(y_1,...,y_n|X_1) > W(y_1,...,y_n|X_2)$ for all $W \in \boldsymbol{W}$ so that the distribution of income for children from disadvantaged groups dominates that of the children from privileged backgrounds. Conversely, the other scenario is the opposite case when $W(y_1,...,y_n|X_1) < W(y_1,...,y_n|X_2)$ for all $W \in \boldsymbol{W}$ where the advantages of being born in privileged families are not wiped out over generations. Both of these scenarios can form bases for claims of intergenerational justice. We take up this point below.

Some theories hold that justice is a social contract without which individual members of society would be worse off. Plato's accounts in the Republic for the existence of justice have been formalized in the second half of this century using bargaining concepts(3). As a result of the bargaining approach, initial conditions of parties (known as status quo points in game theory) were given great importance in the determination of "fair" allocations of resources. If initial conditions are given a role to play in the *fair* division of resources, we may modify our justice criterion in two directions. The first would be to argue that since advantaged groups (in theory) have less to gain from income redistribution than the poor, justice could prevail if the children from privileged backgrounds were better off than those from disadvantaged origins, provided

(3) For a discussion see Barry (1989) ch.2

the difference in living standards of these two groups was not too high.

By taking into account initial conditions, we can argue for an opposite recommendation: that justice prevails only when the children of the poor are made better off than the children from privileged backgrounds. One way to do so would be to appeal to Nozick's (1974) *entitlements* theories. Nozick's concern is not with the allocation of resources reached, but rather with the process with which it is reached. If indeed discrimination and unequal educational opportunities are the mechanisms which are at the root of the inheritance of low income, than we may according to *entitlements* theories, wish to compensate the children of the poor for their disadvantage. In theory nothing precludes us from envisaging a scenario whereby the order of dominance of distributions is reversed in the name of justice. If on the other hand it were ability differentials as opposed to unequal economic opportunities which were at the root of the dominance of income distributions, Nozick's entitlement theories would no longer give us any rationale for redistribution. Under the later scenario, the children from privileged backgrounds are entitled to their superior ability endowments, and the market process is fair. Any redistribution of income would then constitute a violation of rights and would thus be unjust.

In the next section we contrast our proposed criterion of justice with well known measures of intergenerational mobility. These include the regression coefficient between the incomes of parents and children, and the Atkinson (1981) class of measures[4].

### 4.3 Relation to other approaches

To many economists, the concept of justice is taken as synonym for the maximization of social welfare. The origin

---

[4] The discussion that follows is by no means exhaustive, and more complete reviews can be found in Schokkaert and Van de gaer (1993), and Dardanoni (1991). Also see Atkinson et al. (1992) and Bartholomew (1973).

of this association is perhaps due to Mill's conception of justice as a maximizing theory of a utilitarian objective, as well as the subsequent formalizations of Vickrey (1945) and Harsanyi (1953,55). Yet many political philosophers do not endorse the view that the maximization of a social welfare function is an adequate concept of justice. In this context, Ryan (1993) p.5, writes:

"Here it is enough to notice that it is hard to believe that justice is wholly explained by utility, as that it has nothing to do with utility at all."

In a review article on economic justice, Sen (1987) writes:

"...it is fair to say that in traditional welfare economics, when the notion of justice has been invoked, it has typically been seen only as a part of a bigger exercise, viz., that of social welfare maximization rather than taking justice as an idea that commands attention on its own."

The necessity of separating out the concepts of justice and social welfare maximization, has become more apparent with the development of non-utilitarian theories of justice in the last two decades. The work of Rawls (1971) places liberty ahead of aggregate welfare maximization, Nozick (1974) gives priority to entitlements, while Sen (1985) takes capabilities as being the appropriate concept of assessing people's advantage. Our justice criterion does fall in this latter category. It differs from the utilitarian objective of maximizing social welfare, by giving priority to the condition of welfare dominance.

Shorrocks (1978) follows an axiomatic approach to the measurement of mobility. Though his work pertains to mobility transition matrices, it is of relevance to our discussion. To see that this is the case, recall that each line of a transition matrix defines a conditional probability distribution (as discussed in section 4.2). Thus, within our framework if we were to consider c income groups, and if were to group data also over c income groups, our approach would amount to making normative judgements on the basis of a $c \times c$ transition matrix.

Shorrocks considers several axioms, of which we discuss the following four:

(N): Normalization; a mobility index M takes values in the interval [0,1].

(M): Monotonicity; if the probability of exiting a particular state increases, then mobility should be higher.

Two axioms follow from the *normalization* and *monotonicity* axioms:

(I): Immobility; the identity matrix takes the minimum value of the index, $M(I)=0$.

(PM): Perfect Mobility; if for a transition matrix Q, for all i and j the probability $q_{ij}$ is independent of the original position i, $M(Q)=1$.

Shorrocks points out that the axioms may clash since N, M and PM are generally incompatible. However, Geweke et al. (1986) theorem 1, are able to show that for all transition matrices with real non-negative eigen-values the above axioms are mutually consistent.

Let us examine whether our injustice index is consistent with the above axioms. (N) does not raise any problems since the judgements we make are qualitative in nature. Shorrocks' immobility axiom is compatible with (i)of our definition 2.1, since under (i) all income distributions of second generation groups of individuals can be ranked independently of the exact form of u(.). (PM) is the counter-part of (iii) in definition 2.1; where injustice is rejected when the various conditional income distributions are identical. However, the two approaches differ with respect to Shorrocks' monotonicity axiom. (M) does not appear to be sensitive to the direction of mobility from the diagonal; i.e upward and downward mobility a priori seem to receive symmetric treatments.

To clarify this point, write the ith line of a transition matrix Q as $q_i=[\, q_{i1}, \ldots, q_{ii}, \ldots, q_{ic}]$ . Define now the transition matrix Q1 which differs from Q in that the ith line is replaced by

$$q_i^1=[q_{i1}, \ldots, q_{ii-1}, q_{ii}-\varepsilon, q_{ii+1}+\varepsilon, \ldots, q_{ic}]$$

Also define Q2, where the ith line of Q is replaced by

$$q_i^2=[q_{i1}, \ldots, q_{ii-1}-\varepsilon, q_{ii}+\varepsilon, q_{ii+1}, \ldots, q_{ic}]$$

Under the monotonicity axiom of Shorrocks Q1 and Q2 would be more mobile than Q, however Q1 and Q2 do not receive separate treatment with respect to mobility. However, axioms of efficiency preference and inequality aversion would rank $q_i^1$ as dominant over $q_i$, and the latter distribution as dominant over $q_i^2$.

One important class of intergenerational mobility measures is that defined by Atkinson (1981). Atkinson has investigated the class of mobility measures defined over the joint realisations of the parents and children's incomes. Writing the underlying social welfare function as

$$W = \int\int v(x,y)\; f(x,y)\; dx\; dy \qquad (3.1)$$

one is then able to characterize how changes in the joint distribution of parents and children incomes, f(x,y), influences the overall level of welfare. In order to focus entirely on re-ranking, or *exchange mobility*, here we confine our discussion to intergenerational mobility processes where the marginal distributions f(x) and f(y) are identical. Assume initially that v(x,y) is additive:

$$v(x,y) = u_1(x) + u_2(y) \qquad (3.2)$$

the social welfare function W can be then be expressed as :

$$W = \int\int \{ u_1(x) + u_2(y)\}\; f(x,y)dx\; dy$$

or alternatively as,

$$W = \int u_1(x) f(x) dx + \int u_2(y) f(y) dy \qquad (3.3)$$

When v(x,y) is additive, the overall level of welfare is solely determined by the marginal distributions f(x) and f(y); rather than by the extent to which x and y correlate. Atkinson (1981) writes:

"From some standpoints the adoption of an additive function may appear quite acceptable. It may be argued, for example, that we should only be concerned with the distribution at each date and not worry about the movement between income ranges."

Under the additivity assumption, the social valuation of a second generation individual's income y, is independent of his parents income, x. Atkinson then considers the class of social welfare functions for which an increase in intergenerational mobility raises social welfare. This class, v-, is the class of v(x,y) functions such that the cross derivative $\partial^2 v/\partial x \partial y \leq 0$. For $v \in$ v-,an increase in y lowers the social value of a unit increase in x. Thus, the more correlated are x and y, the lower social welfare becomes. The social welfare function is therefore said to exhibit *preference for mobility* when the cross-derivative of v(x,y) is negative[5] .

An intergenerational mobility measure which belongs to the Atkinson v- class, is that proposed by Dardanoni (1991):

$$D = \int \Phi(x) \left[ \int u(y) f(y|x) dy \right] f(x)\ dx \qquad (3.4)$$

where $\Phi(x)$ is any decreasing function of x, so as to capture the view that a marginal welfare gain for children of the less well-off has a greater social value than a corresponding welfare gain for the children of the rich.

Schokkaert and Van de gaer have discussed the similarity between the approach of Dardanoni and the

---

(5) Atkinson makes further refinements of the approach, and introduces another class, v--, of mobility measures.

Atkinson v- class of mobility measures. In Lemma 21 of their paper they show that the measure D, defined in (3.4), is a member of the Atkinson v- class of measures. They criticize the "Atkinson/Dardanoni" approach on the basis that:

"This amounts to an unequal treatment of children of different descent, in which the children of poor parents are rewarded and children of rich parents are punished."

Schokkaert and Van de gaer follow an "opportunities" approach and offer "ex-ante fairness of intergenerational transmission processes" as an alternative to the Atkinson approach.

Last but not least, the regression coefficient, which we denote by $\beta$, is perhaps the most popular tool used for the measurement of intergenerational mobility. Since it is generally taken that a high correlation between the incomes of parents and children is associated with a low degree of justice, we can take minus the regression coefficient to be the appropriate index of intergenerational mobility. From the recent studies of Solon (1992) and Zimmerman (1992) it would appear that the regression coefficient is of the order of 0.3 to 0.5. One special case where the regression coefficient relates to our justice criterion in a straightforward manner, is when the income distributions of the various socio-economic groups are all equal. Under this specific scenario, $f(y|X_i)=f(y|X_j)=f^*$. The child's economic status is then independently distributed from that of his parents, and therefore $\beta=0$.

To pave the way for our empirical applications, we review in the next section some concepts pertaining to the distribution-free statistical inference about Lorenz curves and empirical distribution functions.

**4.4 Implementation**

In seeking conclusions about the justice of an intergenerational mobility process, we will attempt to be as general as possible. Because we are using samples, we need to test the assumption that, say, welfare is higher in socio-economic group A than in group B. Often this may require making assumptions about the population from which we are sampling. But then our conclusions may be invalidated if we assume that the distribution of income is say, log-normal, whereas it happens to be a Pareto distribution. To remove such limitations, we will follow a distribution-free approach for inferential purposes.

Let us assume that our income variable y has a positive mean $\mu$, and a finite variance $\sigma_{yy}$, and that its cumulative distribution function F(.) is continuous and strictly increasing. A natural way of ranking income distributions can be performed on the basis of order statistics. The distribution theory for order statistics is well known so that tests of hypotheses can be systematically derived along the lines of classical statistical inference. The following result shows why this approach will not yield distribution-free conclusions. Let $e_p$ denote the corresponding sample statistic for $\varepsilon_p$. Also define $\varepsilon^{*\prime}=(\varepsilon_1, \varepsilon_2, ...<\varepsilon_{k-1})$(and $e^{*\prime}=(e_1,e_2,...,e_{k-1})$. Then $n^{1/2}(e^*-\varepsilon^*)$ is asymptotically distributed with zero mean and covariance matrix $\Lambda$ :

$$\Lambda = \begin{bmatrix} p_1(1-p_1)/f_1^2 & \cdots & p_1(1-p_{k-1})/f_1 f_{k-1} \\ \vdots & & \vdots \\ p_1(1-p_{k-1})/f_1 f_{k-1} & \cdots & p_{k-1}(1-p_{k-1})/f_{k-1}^2 \end{bmatrix}$$

where $\varepsilon_1 < \varepsilon_2 <...<\varepsilon_{k-1}$ and $f_p$ denotes $f(\varepsilon_p)$ (see Beach and Davidson (1983)). Because the covariance matrix $\Lambda$ is dependent upon f(.), it is impossible to specify it without giving a parametric form for f(.). Therefore, inferences based on order statistics will not be distribution-free.

On the other hand, we can exploit the fact that the sample mean of any distribution will converge to a normal distribution with the same population mean, and a factor (1/n) times the variance of the population.

Looking back to the definition of the Generalized Lorenz curve in (2.3), we can rewrite GLC(p) as follows:

$$\mathrm{GLC}(p) = F(\varepsilon_p) \int_0^{\varepsilon_p} \frac{u\ f(u)\ du}{F(\varepsilon_p)} \qquad (4.1)$$

so that GLC(p) is in fact a multiple of the mean of the conditional distribution of y, upon it being smaller than $\varepsilon_p$. Noting from (2.1) that $F(\varepsilon_p)=p$, we can express (4.1) as:

$$\mathrm{GLC}(p) = p\ \gamma \qquad (4.2)$$

where $\gamma=E(y/y<\varepsilon_p)$. Define $\Theta$ as the vector of Generalized Lorenz curve ordinates: $\Theta=(p_1\gamma_1,\ p_2\gamma_2\ ,\ldots,p_k\gamma_k)$. Let $\hat{\Theta}$ and $\hat{\gamma}$ denote the corresponding sample statistics to $\Theta$ and $\gamma$. Since $\hat{\Theta}$ contains no more than sample means, we can hope that its asymptotic distribution may be derived independently of the distribution of y. Beach and Davidson (1983) show indeed that the asymptotic distribution of $\hat{\Theta}$ is independent of f:

$$\sqrt{n}\,(\hat{\Theta}-\Theta) \rightarrow N(0,\Sigma)$$

where for $i\leq j$ :

$$\sigma_{ij} = p_i\left[\lambda_i^2 + (1-p_i)(\varepsilon_i-\gamma_i)(\varepsilon_j-\gamma_j) + (\varepsilon_i-\gamma_i)(\gamma_j-\gamma_i)\right] \qquad (4.3)$$

and $\lambda_i^2 = \mathrm{var}(y/y<\varepsilon_p)$. Consider now the case where we have an independent random sample for each of two socio-economic groups of respective sizes $n_1$ and $n_2$, and Generalized Lorenz curve ordinates $\widehat{\Theta_1}$ and $\widehat{\Theta_2}$. Then any test of comparison of the two Generalized Lorenz curves can be based on the vector

$$d = \widehat{\Theta_1} - \widehat{\Theta_2} \qquad (4.4)$$

and the covariance matrix

$$\Omega = \widehat{\Sigma}_1 / n_1 + \widehat{\Sigma}_2 / n_2 \qquad (4.5)$$

Likewise, define $\Phi = (\Phi_1, \ldots, \Phi_k)$ as the vector of population Lorenz curve ordinates, and let $\Phi^*$ be the k-1 vector $(\Phi_1, \ldots, \Phi_{k-1})$ . Beach and Davidson also show that $\widehat{\Phi^*}$ (the sample counter-part of $\Phi^*$) is asymptotically normal in the sense that:

$$n^{1/2} \, [ \, \widehat{\Phi^*} - \Phi^* \, ] \rightarrow N(0, V)$$

where for $i \leq j = 1, \ldots, k-1$

$$v_{ij} = \frac{\sigma_{ij}}{\mu^2} + \left(\frac{p_i \gamma_i}{\mu^2}\right)\left(\frac{p_j \gamma_j}{\mu^2}\right) \sigma_{yy} - \left(\frac{p_i \gamma_i}{\mu^3}\right) \sigma_{jk} - \left(\frac{p_j \gamma_j}{\mu^3}\right) \sigma_{ik} \qquad (4.6)$$

where $\mu$ and $\sigma_{yy}$ are respectively the population mean and variance of y.

Finally, let $F_n$ and $G_n$ denote two empirical distribution functions constructed on the basis of two independent samples of size n, drawn from unknown populations F and G. Define the following two vertical distances

$$T_{FG} = n^{1/2} \sup_y \, [F_n(y) - G_n(y)]$$

and (4.7)

$$T_{GF} = n^{1/2} \sup_y \, [G_n(y) - F_n(y)]$$

$T_{FG}$ and $T_{GF}$ are distributed as Kolmogorov-Smirnov (K.S) statistics (see Mood et al. (1974) pp. 508-510), and can be used to test first order stochastic dominance (FOSD) hypotheses (Whitmore (1978, McFadden (1989)). Let $K_{1-\alpha}$ denote the $(100-\alpha)\%$ quantile of the K.S. distribution. Then consider the following two cases:

(C1): $T_{FG} < K_{1-\alpha} < T_{GF}$ (4.8a)

(C1): $T_{GF} < K_{1-\alpha} < T_{FG}$ (4.8b)

If C1 arises, we may conclude that F dominates G in the FOSD sense, at a $\alpha\%$ probability of type I error. Conversely if C2 occurs, we conclude that G dominates F. If neither of C1 or

C2 occurs, we conclude that none of the two distributions is dominant at the specified significance level.

**4.5 Data Limitations**

The observations we will be using in order to illustrate our justice criterion were extracted from wave XX of the University of Michigan's Panel Study of Income Dynamics. The *fathers and sons* sample was described in section 0.2 of our introductory chapter, and its adequacy for the study of intergenerational mobility was discussed in section 0.3 of the same chapter. There we mentioned that the data we ideally required were lifetime profiles on incomes. Instead, as Jenkins(1987) puts it, we possess "snapshot" observations as opposed to complete "movies" on incomes.

Given these data limitations, we have to select the most appropriate indicator variables of economic status. The variables we opt for are the average hourly earnings of fathers and their sons. Other possible indicators include occupational status, family income, and annual earnings. We discuss the strengths and weaknesses of these alternative indicators in turn.

Occupational status does appear to be attractive at first sight since it would appear that it is not subject to measurement error in the way that income and earnings are. There are several disadvantages in using occupational status though. The distribution of workers by occupation is largely influenced by the state of technology. A high level of intergenerational occupational mobility can be recorded because there are hardly any milkmen left, because the automobile industry is highly automatised today, and because there were hardly any computer engineers and information systems experts, twenty years ago. There is also the fact that some children may inherit the occupation of the parents, but may become incompetent in an area where their parents excelled. A high degree of earnings mobility is compatible with little occupational mobility. There is also the other case, where the son of a mediocre plumber becomes an equally incompetent professional.

A high degree of occupational mobility may therefore conceal a high degree of inheritance in economic status. There is also a highly subjective element residing in the classification of jobs into various occupations, and in the ordinal ranking of occupations. This subjective element is much less present in the definition of earnings classes due to the quantitative nature of the underlying variable.

As an alternative to occupational status, we may want to use income. Its quantitative nature makes it a more objective indicator than occupational status. Income is derived from many sources, and may contain non-monetary components such as free housing, food stamps, and other benefits in kind. Earnings, on the other hand, are derived from a unique source, namely the sale of one's labour services. Differences in income may also be due to differing attitudes towards relying on assistance from others (typically relatives, and the state) thus resulting in heterogeneous discounting of the various sources of income. Income is also due to serious measurement error at both tails of the distribution. Wealth income is also largely undeclared.

Annual earnings provide us with another candidate indicator of economic status. We already stated one of their potential advantage over income, namely the unique source from which they are derived. An exception of course is when people derive earnings from several sources, such as a permanent job coupled with occasional independent activities. The limitation of annual earnings resides in the the fact that differences in the variable are the result of two separate components, annual work hours and hourly pay. We could expect the distribution of annual earnings to be more equally distributed than hourly wages, since on the one hand consumption is a normal good at low incomes, and that leisure is (usually) normal, while consumption may not be at high levels of income. In this sense, the use of hourly rather than annual earnings may be seen as a sensible choice.

Hourly earnings are limited by their low accuracy. In the PSID they are constructed as the ratio of annual earnings to work hours. They are thus likely to be subject to two sources of measurement error, annual wages in the numerator, and annual work hours in the denominator. From the PSID cross-validation study of Bound et al. (1989), we know that both sources of measurement error are present in the construction of hourly earnings. The resulting effect is that the correlation (in logs) between reported values and company records, is as low as 0.56 for hourly earnings, while it takes a value of 0.81 for annual earnings. In theory, nonetheless, differences in hourly wages are due to differences in productivity so that there is a clear economic meaning to be attached to this variable (discrimination on the basis of sex, race and social background are obvious departures from the "theory"). On such grounds, hourly earnings may appear to be the most suitable (though far from ideal) choice for an indicator of economic status.

**4.6 Application to the Fathers and Sons Sample**

In our empirical application we compare the earnings distribution of those individuals whose fathers belonged to the lower 50% earnings quantile, with the distribution for those who had fathers in the upper earnings quantile. Earnings of fathers were reported for 1967, while those of sons pertained to 1986. We refer to this analysis as that of "top versus bottom". Our analysis would then be informative of the pattern of intergenerational mobility if reported hourly earnings of parents preserved the same ranking (between lower and upper quantiles) as their life-cycles incomes, and likewise if children's wages preserved an identical ranking as their own life-cycle incomes. These are of course strong assumptions. It is therefore necessary to treat the analysis that follows as being indicative as opposed to confirmative. On such basis, one should also be cautious in drawing any definite conclusions about intergenerational justice in the United States.

Figure 4.2 depicts the Generalized Lorenz curves for the two groups of second generation individuals. The GLC curves are plotted at 10 different points. On the horizontal axis, the pth point pertains to the bottom p/10 fraction of the population. In other words, it is defined as the pth decile. The GLC curve for sons from the top earnings class lies everywhere above that of sons from the bottom earnings class.

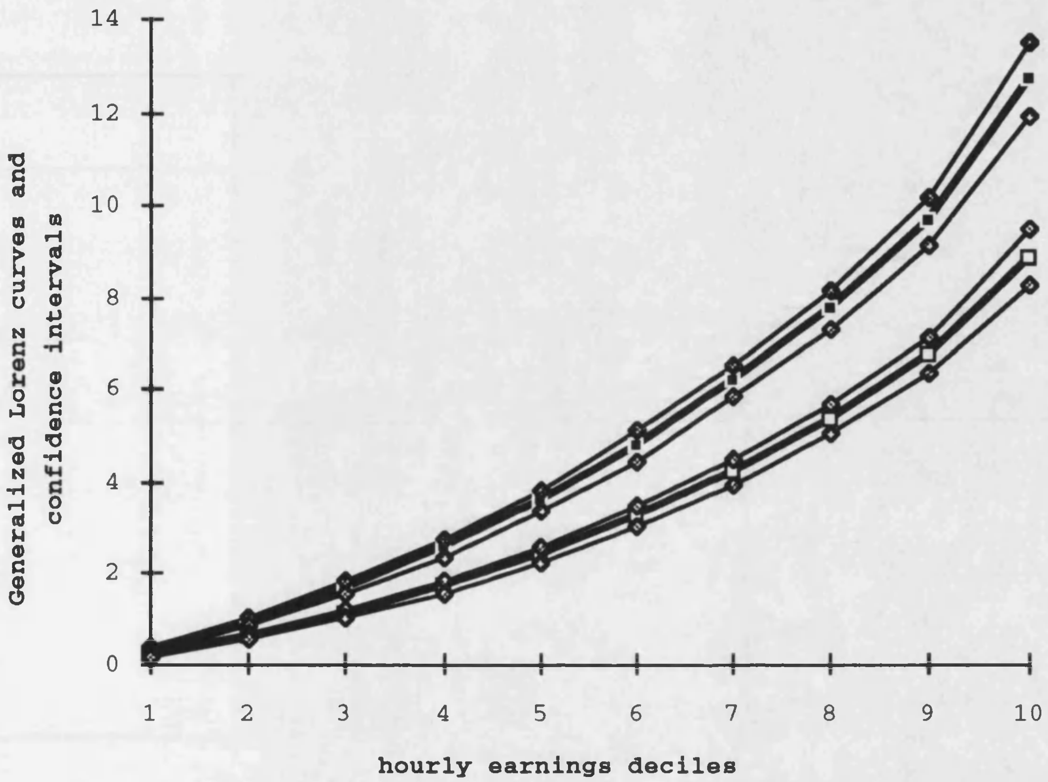


FIGURE 4.2 top 50% versus bottom 50%, Generalized Lorenz curves and 95% confidence intervals

At first sight it therefore appears that the former distribution dominates the latter, so that the studied mobility process exhibits injustice. In order to validate the injustice hypothesis, we need to test the assumption that the GLC curve for the sons from the privileged earnings background lies at every point above that of the other group. The statistic $Z_p$ defined below,

$$Z_p = d_p / \omega_{pp}$$

is asymptotically distributed as a N(0,1) variable under the assumption that the two Generalized Lorenz curves are equal. $\omega_{pp}$ is the pth diagonal element of $\Omega$, and $d_p$ is the difference between the two Generalized Lorenz curves at the pth decile. Table 4.1 below reports the values for the Generalized Lorenz curves ordinates, standard errors in brackets, and the corresponding $Z_p$ statistics.

| p | $GLC_2(p)$ | $GLC_1(p)$ | $Z_p$ |
|---|---|---|---|
| 1 | 0.341 (0.026) | 0.222 (0.017) | 3.786 |
| 2 | 0.926 (0.047) | 0.608 (0.032) | 5.540 |
| 3 | 1.661 (0.067) | 1.094 (0.049) | 6.831 |
| 4 | 2.544 (0.094) | 1.689 (0.065) | 7.484 |
| 5 | 3.564 (0.120) | 2.404 (0.087) | 7.864 |
| 6 | 4.773 (0.160) | 3.230 (0.107) | 8.027 |
| 7 | 6.186 (0.183) | 4.208 (0.135) | 8.699 |
| 8 | 7.760 (0.212) | 5.348 (0.160) | 9.076 |
| 9 | 9.669 (0.253) | 6.746 (0.194) | 9.171 |
| 10 | 12.713 (0.218) | 8.884 (0.285) | 7.834 |

TABLE 4.1: top 50% versus bottom 50%, Generalized Lorenz curve ordinates

$GLC_2$ is the Generalized Lorenz curve for the sons from the top earnings class, while $GLC_1$ is the corresponding curve for the sons originating from the bottom earnings background. As is seen from the last column, the differences between the GLC curve ordinates are all significant at a 99% significance level. Using the evidence from the individual statistics $Z_p$, we conclude that the earnings distribution for the children originating from the bottom earnings class is dominated by that of the children from the top earnings background. We conclude therefore that the mobility process exhibits injustice.

The fathers and sons sample therefore provide sufficient evidence in order to confirm the injustice hypothesis. In explaining the finding that the distribution for children whose fathers were in the top 50% group in 1967 welfare dominates the distribution of the disadvantaged group, we have to address the following questions:

(i) What is the likely effect of considering alternative definitions of family background? Is the above conclusion maintained if we say compare the children originating from the top 20% income group with those whose parents were in the lower earnings quintile in 1967?

(ii) How do the distributions differ in terms of earnings inequality? Are the differences most striking at the tales or in the middle income groups?

(iii) What is the pattern of stochastic dominance of $f(y|X_2)$ over $f(y|X_1)$ ? Is it one of second order stochastic dominance only, or is there a pattern of first order stochastic dominance too?

(iv) How do the results relate to the pattern of intergenerational mobility of $f(x,y)$?

We take up these points in turn.

*(i) Alternative definitions of family background*

What would the Generalized Lorenz curves look like if we were to compare the children originating from the bottom 20% earnings group with those from the top 20% group? Let us refer to these two groups as L20 and U20 respectively. On the one hand the heterogeneity between these two sub-samples is greater in terms of family background, so that to the extent that economic success is influenced by family background, we would expect the previously reached conclusions to be confirmed. On the other hand, since we are working with only 40% of the data, sample variances may be too high to enable us to reject the justice assumption.

As can be seen from table 4.2, $GLC_{U20}$, the Generalized Lorenz curve of the children from the top 20% earnings background, dominates the corresponding curve for the children originating from the bottom 20% earnings class.

| p | $GLC_{U20}(p)$ | $GLC_{L20}(p)$ | $Z_p$ |
|---|---|---|---|
| 1 | 0.428 (0.040) | 0.174 (0.021) | 5.558 |
| 2 | 1.101 (0.074) | 0.491 (0.051) | 6.787 |
| 3 | 1.962 (0.138) | 0.932 (0.075) | 6.552 |
| 4 | 3.036 (0.193) | 1.474 (0.100) | 7.194 |
| 5 | 4.373 (0.259) | 2.108 (0.129) | 7.828 |
| 6 | 5.882 (0.293) | 2.848 (0.153) | 9.179 |
| 7 | 7.503 (0.329) | 3.678 (0.179) | 10.203 |
| 8 | 9.337 (0.383) | 4.708 (0.228) | 10.388 |
| 9 | 11.490 (0.433) | 6.040 (0.304) | 10.303 |
| 10 | 14.502 (0.561) | 8.462 (0.559) | 7.629 |

TABLE 4.2: top 20% versus bottom 20%, Generalized Lorenz curve ordinates

Once again, the data allow us to confirm the injustice assumption at a 99% significance level. In an earlier draft of the chapter we compared the two sub-samples of whites and non-whites. There we also found similar evidence that white children fared much better than their non-whites counter-parts(6).

*(ii) Testing for inequality differences*

A distribution F may be less equally distributed than another one G, but provided F has a sufficiently higher mean income the resulting pattern may be one of welfare dominance of F over G. As can be read from the last line of table 4.1, the mean value of hourly wages was $12.7 per hour in the privileged group and $8.9 per hour in the disadvantaged

(6) This result is in line with the findings of Bound and Freeman (1992). Borjas (1992) explains the relationship between economic success and cultural background using an *ethnic capital* theory.

group. The $Z_p$ statistic of 7.83 indicates that the difference in average wages is statistically significant from zero. It is therefore interesting to inquire as to which of the two earnings distributions is more equally distributed.

| p | $LC_2(p)$ | $LC_1(p)$ | $Z_p$ |
|---|---|---|---|
| 1 | 0.026 (0.002) | 0.025 (0.002) | 0.655 |
| 2 | 0.072 (0.004) | 0.068 (0.003) | 0.863 |
| 3 | 0.130 (0.005) | 0.123 (0.005) | 1.046 |
| 4 | 0.200 (0.007) | 0.189 (0.006) | 1.170 |
| 5 | 0.280 (0.008) | 0.271 (0.008) | 0.820 |
| 6 | 0.374 (0.010) | 0.363 (0.010) | 0.810 |
| 7 | 0.486 (0.011) | 0.473 (0.011) | 0.850 |
| 8 | 0.610 (0.013) | 0.601 (0.012) | 0.469 |
| 9 | 0.759 (0.014) | 0.758 (0.013) | 0.078 |
| 10 | 1 | 1 | - |

TABLE 4.3: top 50% versus bottom 50%, Lorenz curve ordinates

Though it appears that the distribution of earnings for the top 50% group is more equally distributed ($LC_2(p) \geq LC_1(p)$ for all p), none of the differences in Lorenz curve ordinates are statistically significant. This result points to the conclusion that earnings inequality does not appear to vary across the two socio-economic groups.

*(iii) The pattern of stochastic dominance*

Given that the two earnings distributions appear to be equally distributed, but that $F_2$ (the distribution of the children originating from the top 50% group) has a higher sample mean than $F_1$ (the corresponding distribution for the the bottom 50% group), we are lead to believe that the violation of justice takes the form of a first order stochastic dominance of $F_2$ over $F_1$. Let us look at the transition matrix underlying the process:

$$\Pi = \begin{bmatrix} 0.140 & 0.142 & 0.123 & 0.102 & 0.114 & 0.093 & 0.108 & 0.053 & 0.081 & 0.045 \\ 0.059 & 0.059 & 0.076 & 0.099 & 0.085 & 0.108 & 0.090 & 0.148 & 0.118 & 0.156 \end{bmatrix}$$

By cumulating the elements along the rows one notes that the sum is always smaller along the bottom line. Our initial guess would indeed be that $F_2$ dominates $F_1$.

Let $F_{in}$ denote the sample analogue of $F_i$. A general test of first order stochastic dominance can be based on the Kolmogorov-Smirnov statistic. From our sample

$$\sup_y \left[F_{1n}(y) - F_{2n}(y)\right] = 0.2465,$$

while

$$\sup_y \left[F_{2n}(y) - F_{1n}(y)\right] = 0$$

For n=472, we therefore have that

$$n^{1/2} \sup_y \left[F_{1n}(y) - F_{2n}(y)\right] = 5.355$$

The critical value of the K.S. test at $\alpha$=5% is equal to 1.36. It follows that the largest vertical distance between the two empirical distributions is significant. On the basis of equations (4.8) we may conclude that the earnings distribution $F_2$ for the privileged children first order dominates $F_1$. The implication of this result is that the confirmation of the injustice hypothesis would still be maintained even if we were to drop the assumption that the welfare functions exhibit social aversion to inequality.

*(iv) Explanations based on the pattern of intergenerational mobility*

In order to relate the welfare dominance results previously reached to the pattern of intergenerational mobility, we take a close look at the square (c x c) transition matrix of earnings. The construction of a square size transition matrix is particularly useful in the context of addressing question (iv) above, since it provides a natural set-up for examining the frequency of immobility (the diagonal $\pi_{ii}$ elements), as well as the nature of mobility (viz. the off-diagonal $\pi_{ij}$ elements)

If we divide the sample of fathers and sons into a 5x5 quintile classification (so as to obtain a square matrix with constant row and column marginals), we obtain the following matrix of counts:

$$T = \begin{bmatrix} 63 & 50 & 31 & 24 & 21 \\ 58 & 40 & 41 & 27 & 23 \\ 27 & 32 & 55 & 42 & 33 \\ 26 & 40 & 37 & 56 & 30 \\ 15 & 27 & 25 & 40 & 82 \end{bmatrix}$$

The joint probability matrix of incomes, i.e. the discrete analogue of f(x,y), can be constructed by dividing each element by the total number of observations (n=945). Likewise, $\pi_{ij}=5.p_{ij}=(5/945)t_{ij}$[7].

$$\Pi = \begin{bmatrix} 0.3333 & 0.2646 & 0.1640 & 0.1270 & 0.1111 \\ 0.3069 & 0.2116 & 0.2169 & 0.1429 & 0.1217 \\ 0.1429 & 0.1693 & 0.2910 & 0.2222 & 0.1746 \\ 0.1376 & 0.2116 & 0.1958 & 0.2963 & 0.1587 \\ 0.0794 & 0.1429 & 0.1323 & 0.2116 & 0.4339 \end{bmatrix}$$

The matrices T and $\Pi$ reveal two interesting patterns. Firstly, there appears to be a high degree of inheritance of income status: with the exception of the second row, in each line the diagonal element has the largest fraction of families. This phenomenon is especially marked at the bottom and top lines of the transition matrix. Also, the data tend to reveal a pattern of symmetry. Under the symmetry assumption, the probability of observing upward moves from class i to j, $p_{ij}$, equals the probability $p_{ji}$ of observing a similar downward move. Given the classification of parents and children into c groups of equal size, $\pi_{ij}=cp_{ij}$ , so that symmetry can be stated either in terms of the transition matrix $\Pi$, or the joint distribution matrix P.

A test of symmetry based on the matrix of counts T is described in Everitt (1977) p.155. The statistic $X^2$:

---

[7] More precisely, the transition matrix can be estimated using the Maximum Likelihood method by dividing each entry in a given line by the sum of the line elements (see Anderson and Goodman (1957)).

$$X^2 = \sum_{i<j} \{ (t_{ij} - t_{ji})^2 / (t_{ij} + t_{ji}) \}$$

is asymptotically distributed as a Chi-Square variable with $v=c(c-1)/2$ degrees of freedom, where c is the number of classes. Here $c=5$, so that we have $v=10$ degrees of freedom. From the matrix T above, we calculate $X^2=8.749$. The critical value of our test at a 5 % probability of type-I error, is 18.307. The hypothesis of symmetry thus cannot be rejected. In order to be assured that this result is not purely due to the choice of five classes, we report similar results in the table below for other values of c.

| c | $X^2$ | $\chi^2_{v;0.95}$ |
|---|---|---|
| 3 | 1.516 | 7.815 |
| 4 | 0.559 | 12.60 |
| 5 | 8.749 | 18.307 |
| 6 | 12.149 | 25.0 |
| 7 | 21.311 | 32.7 |
| 8 | 31.21 | 41.337 |

TABLE 4.4: Testing symmetry

We conclude from the calculations reported in table 4.5, that the assumption of symmetry is not rejected (for most values of c), at a 5% probability of type-I error.

The symmetry model therefore provides us with an adequate parametrization of our intergenerational data. It also helps us to understand why these data conform to the injustice scenario: the son of a low income father has a small chance of climbing up to the top earnings group, and likewise the son of a rich man is equally unlikely to experience a similar downward fall. For instance, under the symmetry model, $\pi_{15}=\pi_{51}$, estimated at $(21+15)/(2 \times 189) = 0.096$.

It should be noted that it is not the property of symmetry on its own which causes the confirmation of the injustice scenario. To take a limiting case, if $\pi_{ij}=1/c$ for all i and j, the intergenerational mobility process is symmetric, and justice obtains (as a result of equality of distributions). What explains the welfare dominance pattern in favour of the privileged children, is the fact that the diagonal elements of $\Pi$ are fairly large in magnitude. For example, $\pi_{11}$ is estimated at 1/3, and $\pi_{55}$ at 0.43. It therefore appears that it is the combined facts of symmetry coupled with lack of mobility which taken together, account for the confirmation of the injustice scenario.

**4.6 Conclusions**

We have suggested that the justice assessment of intergenerational mobility processes be carried out via the comparison of income distributions of the second generation families originating from various socio-economic groups. The principle guiding our assessment was one of welfare dominance, meaning that justice is rejected when it can be concluded that one socio-economic group is unambiguously better off than any other one.

Bearing in mind that life-cycle biases, measurement errors, and lack of population representativeness, are all serious sources of data limitations, we conclude that the intergenerational data we have examined lead to the confirmation of the injustice scenario. This result is explained by noting the occurrence of an apparently symmetric pattern of movement between income classes, coupled with relatively high income inheritance probabilities. The resulting effect is such that the income distributions of children originating from privileged backgrounds dominate the distributions for children from less favourable economic backgrounds.

These conclusions were reached on the basis of the PSID. Other U.S. data sets ought to be examined before one makes strong assertions about the pattern of intergenerational mobility in the United States. Also, such findings need not be valid for other countries. The study of Erikson and Goldthorpe (1992) suggests that amongst developed economies some differences in the pattern of intergenerational mobility can be observed between on the one hand the group of European nations, and on the other hand, the U.S. Australia and Japan. Socialist economies may be expected to exhibit a different pattern from market economies. The stage of economic development may also be an important factor. The time horizon over which incomes are observed could also be an important element. Studies of occupational mobility suggest that individuals experience some downward mobility in their first years of work, but eventually converge to their parents occupational group. Finally we note that our sample only looked at the mobility of men. Because the labour force participation of women is governed by different circumstances from that of men, there is no reason to expect the patterns of intergenerational mobility of men and women to be similar.

CHAPTER 5

## Conclusions and Directions for Future Research

### 5.1 Summing Up

In the first two chapters of the thesis we have dealt with the questions of identifying the poor and allocating benefits to individuals in a situation of income uncertainty. We suggested to identify the poor using multiple indicators of living standards. From a p x 1 vector of indicators, a uni-dimensional sufficient statistic was constructed in order to rank families in the permanent income space. At a practical level, the multiple indicator index appeared to correlate highest with all three indicators considered in our empirical applications. This observation lead us to conclude that the suggested framework was a sensible compromise between ranking families according to food budget share exclusively, versus using an income definition in the identification of the poor.

Using the framework developed in chapter one, we have then addressed in chapter two the question of the allocation of state benefits in a situation of income uncertainty. We have adopted a decision-theoretic approach to the problem. The solution then consisted in granting assistance to a family if the expected social benefit from the transfer outweighed the corresponding social cost. Increases in the size of the transfer and in the social opportunity cost of poverty alleviation policies reduced the likelihood that a given family qualifies for state assistance, while rises in the poverty line made it more likely.

The purpose of our empirical section was to illustrate how changes in the parameters of the problem altered the degree of coverage of a hypothetical poverty alleviation programme. As we moved from restricted coverage towards a universal benefit system, the pool of recipients revealed on average better incomes and employment possibilities, and lower food budget shares. In this sense, the decision theoretic framework adopted in the allocation of benefits under uncertainty seemed to do the job adequately.

In chapter 3 we examined the relationship between credit market imperfections and the incidence of poverty. Our main finding was that credit market imperfections did not increase the incidence of poverty, and furthermore could allow poverty reductions. This result was obtained in a setting of adverse selection. The conclusions depend to a certain extent on the assumptions underlying the functioning of the credit market. Nonetheless the findings are useful in illustrating two points.

Firstly we note that the distinction between *first* and *second best* economies becomes misleading when we consider other objectives than the maximisation of social surplus. When poverty reduction is taken as the social objective, the economy operating under a regime of imperfect information (a typical second best case) may out-perform the first best economy in terms of achieving a lower incidence of poverty. The second point the model serves to illustrate, is the importance of assessing the likely impacts of efficiency enhancing policies on the incidence of poverty. As was pointed out in a different context by Atkinson (1993b), it may well be that government regulatory policy in an imperfect market has an adverse effect on the level of poverty.

The purpose of chapter 4 was to suggest a criterion for the justice assessment of intergenerational mobility processes. Our interest in the question was motivated by the empirical evidence as well as the theoretical models that establish the importance of family background in the determination of income. We were thus lead to qualify an intergenerational mobility process as being unjust if it were possible to obtain a welfare ranking of income distributions for any two groups of individuals (using a classification based on socio-economic background). The evidence from our *fathers and sons* sample pointed towards the injustice scenario.

This conclusion however cannot be taken at face value. Given the many data shortcomings involved with studies of this nature, one can only treat the results as suggestive of a particular scenario. Further research is required in order to take into account sample selection and attrition biases, as well as the measurement error problem. We take up these points in the section below, where we discuss directions for further research on the basis of what has been reached in the dissertation.

**5.2 An Agenda for Future Work**

In our first chapter we have brought together Friedman's theory of the consumption function and the method of factor analysis, in order to address the question of the identification of the poor. As discussed in Deaton (1992), empirical tests of the permanent income hypothesis often lead to its rejection. It may therefore be interesting to consider the question of the identification of the poor on the basis of other theories of consumption. We have also mentioned in chapter one that as an alternative to factor analysis one could use other multivariate data reduction methods such as principal component analysis.

Within factor-analytic methods, one could also consider more sophisticated models and estimation methods. Our model of chapter one was just-identified, but there may be reasons for fitting over-identified models, in which case estimation methods such as ordinary least squares, generalized least squares, or maximum likelihood, may provide different estimates. In the case of just-identified models, all three estimation methods are equivalent, so that in chapter one we did not have to worry about the choice of estimation technique. The choice of indicators is also an important question. Theoretical considerations may guide our initial choices, but data limitations may lead us to search for other variables. There may be theoretical grounds for selecting a set of indicators X* over X, however the fitted covariance matrix for the error vector may fail to be positive definite using X* (1), but not using X.

We have mentioned in chapter two the many factors we have omitted in our analysis of the allocation of benefits under uncertainty. These included the disincentive effects of state transfers, the take-up problem, as well as transfer costs. The framework could be further developed in order to make the size of the transfer a choice variable, as in Glewwe (1990).

Given the current policy debates about the choice between means-testing, versus universal provision of benefits, we feel that the model could be formulated in different terms in order to make it appear more relevant to policy analysis. The relationship between income and the poverty line may for instance be modelled as a productivity relationship, as opposed to solely being based on welfare considerations. Type-I errors would then be seen as resulting in productivity losses, not just depriving the poor from enjoying higher welfare levels. We are too often reminded of the disincentive effects of transfers. However, we feel that the incentive effects of transfers may be far more important for the poorer groups of society, especially within the context of developing countries. Nutritional models (for e.g. Dasgupta and Ray (1986)), efficiency wage arguments (Weiss (1991)), and also our model of chapter 3, point towards the argument that state transfers may result in efficiency gains.

As illustrated in our empirical applications, increases in the poverty line entail a higher degree of benefit coverage. One could then use these results to illustrate that say, even if the poverty line is set at the 10% quantile, the next 15% from the population may also derive utility from public transfers. This would be one way of modelling the effect of poverty alleviation policies on the *near-poor*. Similar extensions of the present framework

---

(1) In the factor analysis literature, such situations are referred to as *Heywood cases*.

could make it appear more in line with the current policy debate about targeting versus universal provision of benefits.

One weakness of the model of chapter 3 can be seen by contrasting its predictions about the pattern of intergenerational earnings mobility, with the empirical findings of chapter 4. The model does not have any downward mobility, whereas the intergenerational transition matrix of earnings exhibits a symmetry pattern. How do we reconcile these two facts?

One attitude would consist of rejecting the findings, since sample selection biases and measurement errors in earnings are too serious for us to have any faith in our results. We return to this point below. Another way of reconciling data and theory, would be to calibre the predictions of the model of chapter 3 towards the symmetry pattern. In a previous version of the model, both funded and non-funded agents came as high and low return to education individuals. The economy thus consisted of four types of workers: NH, NL, FH, and FL. Under the assumption that the distributions of ability and wealth were independent, it was shown that both low return types (NL and FL) had identical reservation wages despite the differences in their wealth backgrounds. Then we can show that when the demand for skilled labour is not high enough to induce the participation of the low returns individuals, but sufficiently high for NH-type individuals to wish to acquire education, some downward mobility may be observed. By this we mean that FL individuals may end up earning lower wages than NH individuals. Though such a model would entail some downward mobility, it does not follow that the resulting overall pattern of movement between classes may be symmetric. Piketty's (1992) model also entails some downward mobility, but symmetry does not follow from the predictions of his model either.

Our feeling is that there is more to the symmetry result than just noisy data. We would thus be inclined to conclude that it is a question worth pursuing at the theory level, and that explanations may have to be sought through other mechanisms than the relationship between the allocation of credit and the the distribution of income. Wealth and ability backgrounds are of course important determinants of income, but as discussed by Bowles(1973), there may be other equally important cultural and social factors that influence the overall pattern of intergenerational mobility.

Data limitations are the major problem of empirical studies on intergenerational mobility. For this reason, little can be concluded about the findings of chapter 4. More work is required in modelling the probability processes underlying intergenerational samples. Methods of correcting for sample selection biases are well developed in econometrics (for e.g. Amemiya (1985) ch. 10). As pointed out by Hsiao (1986) and Atkinson et al.(1992), these same methods can be used to correct for attrition biases in longitudinal data sets.

Errors in variables methods are widely used in regression analysis. Typical applications in empirical labour economics consist of modelling the determinants of earnings. While the "purging" of data from various sources of "contamination" initially appeared an attractive solution, it is now apparent from cross-validation studies on incomes that the underlying assumptions in the errors in variables methods were wrong(2). Great care is therefore needed in order to avoid further distorting reality with good intentions.

Nonetheless, the fact remains that hourly earnings are imperfect indicators of long-run economic status, just like current income is a noisy indicator of permanent income. The techniques of chapter 1 may therefore be used in order to rank fathers and sons according to long-run income status. Whether these predicted long term incomes are biased will not be crucial if we decide to work with transition matrices. Then what would matter would be the ranks of

(2) See Heckman (1993) for a critical survey on empirical labour economics.

individuals, as opposed to the distances between their incomes.

Alternatively we may choose to work with transition matrices directly taking into account measurement errors[3] :

$$\underset{\text{probability}}{\text{sample transition}} = \underset{\text{probability}}{\text{population transition}} + \text{error}$$

Parametric models also need to be elaborated in order to further acquire understanding about the properties of intergenerational mobility processes. Abul Naga and Antille (work in progress) propose the model

$$\Pi = A\,S + (I-A) + E$$

in the analysis of square intergenerational transition matrices. A is a diagonal matrix of unknown parameters in the interval $0 \leq a_{ii} \leq 1$, S is an unknown symmetric transition matrix, and E is a matrix of errors. More general models are also required in order to nest specific scenarios of interest in the field of intergenerational mobility. Such developments would further deepen our understanding of an important question in the social sciences.

---

(3) See chapter 11 of Amemiya (1985) for a discussion.

REFERENCES

Abul Naga R.(1993): "Testing the Symmetry of the Intergenerational Matrix of Earnings", mimeo, LSE.

Aghion P. and P. Bolton (1991): "A Trickle-Down Theory of Growth and Development with Debt-Overhang", mimeo, LSE.

Agresti A. (1984): *Analysis of Ordinal Categorical Data*, New York, John Wiley.

Akerlof, G. (1970): " The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism", *Quarterly Journal of Economics* vol. 84, 488-500.

Akerlof, G. (1978): "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning", *American Economic Review* vol. 68, 8-19.

Amemiya T. (1985): *Advanced Econometrics*, New York, Blackwell.

Anand S. and C. Harris (1991): "Food and Standard of Living: An Analysis Based on Sri Lankan Data", in Dreze J. and A. K. Sen (eds.): *The Political Economy of Hunger*, Oxford, Clarendon Press.

Anderson T. W. and L. Goodman (1957): "Statistical Inference About Markov Chains", *Annals of Mathematical Statistics* vol. 28, 89-110.

Anderson T.W. (1984): *Multivariate Analysis*, New York, John Wiley.

Arrow K. (1973): "Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice", *Journal of Philosophy* vol. 70, 245-263.

Atkinson A. B. (1981): "The Measurement of Economic Mobility", in *Essays in Honour of Jan Pen*, reprinted in Atkinson (1983b).

Atkinson A. B. (1983a): "Intergenerational Earnings Mobility in Britain", in *Ansatze der Lebenseinkommensanalyse, Herausgegeben von Winfried Schmahl*, Tubingen, J.C.B. Mohr.

Atkinson A. B. (1983b): *Social Justice and Public Policy,* London, Harvester Wheatsheaf.

Atkinson A. B. (1989): *Poverty and Social Security*, London, Harvester Wheatsheaf.

Atkinson A. B.(1993a): "On Targeting Social Security: Theory and Western Experience with Family Benefits", Welfare State Programme Discussion Paper no. 99, London School of Economics.

Atkinson A. B. (1993b): "Capabilities, Exclusion, and the Supply of Goods", Welfare State Programme Discussion Paper no. 97, London School of Economics.

Atkinson A. B., A. K. Maynard, and C. Trinder (1983) : *Parents and Children : Incomes in two Generations,* London, Heinemann.

Atkinson A. B., F. Bourguignon, and C. Morisson (1992) : *Empirical Studies of Earnings Mobility*, Paris, Harwood Academic Publishers.

Barham V., R. Boadway, M. Marchand, and P. Pestieau (1992): "Education and the Poverty Trap", *C.O.R.E. discussion paper* no. 9210, Universite Catholique de Louvain.

Barnett V.(1982): *Comparative Statistical Inference*, New York, John Wiley.

Barry B. (1989): *Theories of Justice*, London, Harvester Wheatsheaf.

Bartholomew D. J. (1973) : *Stochastic Models for Social Processes*, London, John Wiley.

Bartholomew D. J. (1981): "Posterior Analysis of the Factor Model", *British Journal of Mathematical and Statistical Psychology* vol. 34, 93-99.

Bartholomew D. J. (1984): "The Foundations of Factor Analysis", *Biometrika* vol. 71, 221-32.

Bartholomew D. J. (1985): "Foundations of Factor Analysis: Some Practical Implications", *British Journal of Mathematical and Statistical Psychology* vol. 38, 1-10.

Bartholomew D. J. (1987): *Latent Variable Models and Factor Analysis,* New York, Oxford University Press.

Beach C. and R. Davidson (1983): "Distribution-Free Statistical Inference with Lorenz Curves and Income Shares", *Review of Economic Studies*, L, 723-735.

Becker G. (1981): *A Treatise on the Family*, Cambridge MA, Harvard University Press.

Becker G. and N. Tomes (1986): "Human Capital and the Rise and Fall of Families", *Journal of Labor Economics* vol. 43, PS1-PS39.

Becketti S., W. Gould, L. Lillard , and F. Welch (1988): "The Panel Study of Income Dynamics After Fourteen Years: an Evaluation", *Journal of Labor Economics* vol. 6, 472-92.

Behrman J. and P. Taubman (1985): "Intergenerational Mobility in the U.S.: Some Estimates and a Test of Becker's Intergenerational

Endowments Model", *Review of Economics and Statistics* vol. 67, 144-51

Behrman J. and P. Taubman (1990): "The Intergenerational Correlation Between Children's Adult Earnings and Their Parents Income: Results from the Michigan Panel Study of Income Dynamics", *Review of Income and Wealth* vol. 36, 115-27.

Bernanke B. and G. Gertler (1990): "Financial Fragility and Economic Performance", *Quarterly Journal of Economics* vol. 105, 87-114.

Besley T. (1990): "Means Testing versus Universal Provision in Poverty Alleviation Programmes", *Economica* vol. 57, 249-261.

Besley T. and S. Coate (1992): "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs", *American Economic Review* vol. 82, 249-61.

Bishop J., S. Chakraborti, and P. Thistle (1989): "Asymptotically Distribution-Free Statistical Inference for Generalized Lorenz Curves", *Review of Economics and Statistics* vol. 71, 725-727.

Bishop Y., S. Fienberg, and P. Holland (1975): *Discrete Multivariate Analysis: Theory and Practice,* Cambridge MA, M.I.T. Press.

Bollen K. (1990): *Structural Equations with Latent Variables*, New-York, John Wiley.

Borjas G. J. (1992): "Ethnic Capital and Intergenerational Mobility", *Quarterly Journal of Economics* vol. 107, 123-50.

Bound J. and R. B. Freeman (1992): "What Went Wrong? The Erosion of Relative Earnings and Employment Among Young Black Men in the 1980s", *Quarterly Journal of Economics* vol. 107, 201-232.

Bound J., C. Brown, G. Duncan, and W. Rodgers (1989): "Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Results From Two Validation Studies", NBER Working Paper no. 2884, Cambridge MA.

Bowles S. (1973): "Understanding Unequal Economic Opportunity", *American Economic Review* vol. 63, 346-356.

Chaudhuri S. and M. Ravallion (1992): "How Well Do Static Indicators Identify the Poor?", mimeo, World Bank.

Chiang A. (1984): *Fundamental Methods of Mathematical Economics*, London, McGraw-Hill.

Conlisk J. (1974): "Can Equalization of Opportunity Reduce Social Mobility?", *American Economic Review,* vol. 64, 80-90.

Conlisk J. (1984): "Four Invalid Propositions About Equality, Efficiency, and Intergenerational Transfers Through Schooling", *Journal of Human Resources* vol. 19, 3-21.

Dardanoni V.(1991): "Measuring Social Mobility", University of California at San Diego, discussion paper.

Dasgupta P. and D. Ray(1986): "Inequality as a Determinant of Malnutrition and Unemployment, Theory", *Economic Journal* vol. 96, 1011-34.

Deaton A.(1992): *Understanding Consumption*, Oxford, Clarendon Press.

Deaton A. and J. Muellbauer (1980): *Economics and Consumer Behaviour*, Cambridge, Cambridge University Press.

Dixit A. (1990): *Optimization in Economic Theory*, Oxford, Oxford University Press.

Dreze J. and A. K. Sen (1990): *Hunger and Public Action*, Oxford, Clarendon Press.

Dreze J., P. Lanjouw, and N. Stern (1992) : "Economic Mobility and Agricultural Labour in Rural India: a Case Study", D.E.P. discussion paper no. 35, Sticerd, London School of Economics.

Duclos J.(1992): "Poverty Alleviation and Redistibutive Costs", mimeo.

Eatwell J., M. Milgate, and P. Newman (eds.) (1987): *The New Palgrave: A Dictionary of Economics*, London, Macmillan.

Erikson R. and J. Goldthorpe (1992): *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford, Clarendon Press.

Everitt B.S. (1977): *The Analysis of Contingency Tables*, London, Chapman and Hall.

Everitt B.S. (1984): *An Introduction to Latent Variable Models*, London, Chapman and Hall.

Feller W. (1968): *An Introduction to Probability Theory and Its Applications*, New York, John Wiley.

Friedman M. (1957): *A Theory of the Consumption Function*, Princeton, Princeton University Press.

Friedman M. and S. Kuznets (1945): *Income From Independent Professional Practice*, New York, National Bureau of Economic Research.

Galor O. and J. Zeira (1993): "Income Distribution and Macroeconomics", *Review of Economic Studies* vol. 60, 35-52.

Geweke J. , R. Marshall, and G. Zarkin (1986): "Mobility Indices in Continuous Time Markov Chains", *Econometrica* vol. 62, 1407-1423.

Glewwe P. (1990): "Efficient Allocation of Transfers to the Poor", LSMS working paper no. 70, Washington D.C., World Bank.

Glewwe P. and J. Van Der Gaag (1990): "Identifying the Poor in Developing Countries: Do Different Definitions Matter?", *World Development* vol. 18, 803-814.

Goldberger A. (1972): "Structural Equations Methods in the Social Sciences", *Econometrica* vol. 40, 979-1001.

Goldberger A. S. (1989): "Economic and Mechanical Models of Intergenerational Transmission", *American Economic Review* vol. 79, 504-513.

Goodin R. (1985): "Erring on the Side of Kindness in Social Welfare Policy", *Policy Sciences* vol. 18, 141-156.

Greene W.(1991): *Econometric Analysis*, New York, MacMillan.

Hammond P. (1987): "Envy", in Eatwell et al.

Hanratty M. J. and R. M. Blank (1992): "Down and Out in North America: Recent Trends in Poverty Rates in the United States and Canada", *Quarterly Journal of Economics* vol. 107, 233-254.

Harsanyi J. (1953): "Cardinal Utility in Welfare Economics and the Theory of Risk-Taking", *Journal of Political Economy* vol. 61, 434-5.

Harsanyi J. (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", *Journal of Political Economy* vol. 63, 309-21.

Hayek F.A. (1982): "Social or Redistributive Justice" , from *Law, Legislation and Liberty*, reprinted in Ryan (1993).

Heckman J. (1993): "What Has Bean Learned About Labor Supply in the Past Twenty Years?", *American Economic Review* vol. 83, 116-121.

Hirshleifer J. and J. Riley (1992): *The Analytics of Uncertainty and Information,* Cambridge, Cambridge University Press.

Hoff K. (1991): "The Second Theorem of the Second Best", mimeo, University of Maryland.

Hsiao C. (1986): *The Analysis of Panel Data*, Cambridge, Cambridge University Press.

Jenkins S. (1987): "Snapshots Versus Movies, 'Lifecycle Biases' and the Estimation of Intergenerational Earnings Inheritance", *European Economic Review* vol. 31, 1149-1158.

Kakwani N. (1990): "Large Sample Distribution of Several Inequality Measures with Application to Cote d'Ivoire", LSMS Working Paper no. 61, Washington D.C., World Bank.

Kanbur R. (1987): "Measurement and Alleviation of Poverty", *IMF Staff Papers* vol. 34, 60-85.

Laffont J. (1989): *Fundamentals of Public Economics*, Cambridge MA., M.I.T. Press.

Lillard L. and R. Willis (1978): "Dynamic Aspects of Earnings Mobility", *Econometrica* vol. 46, 985-1012.

Loury G. (1981): "Intergenerational Transfers and the Distribution of Earnings", *Econometrica* vol. 49, 843-867.

Mankiw N. G. (1986): "The Allocation of Credit and Financial Collapse", *Quarterly Journal of Economics* vol. 101, 455-470.

McFadden D. (1989): "Testing for Stochastic Dominance", in Fomby T. and T. Seo (eds.): *Studies in the Economics of Uncertainty in Honour of Josef Hadar*, New York, Springer-Verlag.

McGregor P. P. L. and V. K. Borooah (1992): "Is Low Spending or Low Income a Better Indicator of Whether or Not a Household is Poor? Some Evidence from the 1985 Family Expenditure Survey", *Journal of Social Policy* vol. 22, 53-69.

Meade J. (1973): *The Inheritance of Inequalities*, London, George Allen and Unwin.

Meier G. (1984): *Leading Issues in Economic Development*, Oxford, Oxford University Press.

Mincer J. (1979): "Human Capital and Earnings", in D. Windhan (ed.) *Economic Dimensions of Education*, New York, National Academy of Education.

Mood M., F. Graybill, and D. Boes (1974) : *Introduction to the Theory of Statistics*, Auckland, McGraw-Hill.

Morrison D. F. (1978): *Multivariate Statistical Methods*, Auckland, McGraw-Hill.

Nozick R. (1974): *Anarchy, State and Utopia*, Oxford, Blackwell.

Quah D. (1993): "Galton's Fallacy and Tests of the Convergence Hypothesis", *Scandinavian Journal of Economics* vol. 95, 427-443.

Quah D. (1994): "Convergence Empirics Across Economies with (Some) Capital Mobility", Sticerd Discussion Paper EM/94/275, London School of Economics

Piketty T. (1992): "Imperfect Capital Markets and Persistence of Initial Wealth Inequalities", mimeo, L.S.E.

Ravallion M. (1988) : "Expected Poverty Under Risk Induced Welfare Variability", *Economic Journal* vol. 98, 1171-1182.

Rawls J. (1971): *A Theory of Justice*, Cambridge MA., Harvard University Press.

Riley J. (1976) : "Information, Screening, and Human Capital", *American Economic Review* vol. 66, 254-60.

Ryan A. (ed.) (1993): *Justice*, Oxford, Oxford University Press.

Sawhill I. (1988) "Poverty in the U.S., Why Is It So Persistent?", *Journal of Economic Literature* vol. 26, 1073-1119.

Schokkaert E. and D. Van de gaer (1993): "Ex-ante Fairness of Intergenerational Transmission Processes", mimeo, Centre for Economic Studies, K.U. Leuven.

Schultz T. (1993): *The Economics of Being Poor*, Oxford, Blackwell.

Sen A. K. (1973): "On Ignorance and Equal Distribution", *American Economic Review* vol. 63, 1022-24.

Sen A. K. (1981): *Poverty and Famines, an Essay on Entitlement and Deprivation*, Oxford, Clarendon Press.

Sen A. K. (1985): *Commodities and Capabilities*, Amsterdam, North Holland.

Sen A. K. (1987): "Justice", in Eatwell et al.

Sen A. K. (1992): "The Political Economy of Targeting", mimeo.

Sen A. K. and B. Williams (eds.) (1982): *Utilitarianism and Beyond*, Cambridge, Cambridge University Press.

Shah A. (1992): *Credit Markets and the Distribution of Income*, London, Academic Press.

Shapiro C. and J. Stiglitz (1984): "Unemployment as a Worker Discipline Device", *American Economic Review* vol.74, 433-44.

Shorrocks A. (1978): "The Measurement of Mobility", *Econometrica* vol. 46, 1013-24.

Shorrocks A. (1983): "Ranking Income Distributions", *Economica* vol. 50, 3-17.

Silverman B. (1986): *Density Estimation for Statistics and Data Analysis* , New York, Chapman and Hall.

Small K. and H. Rosen (1981): "Applied Welfare Economics with Discrete Choice Models", *Econometrica* vol. 49, 105-130.

Solon G. (1992): "Intergenerational Income Mobility in the United States", *American Economic Review* vol. 82, 393-408.

Spence M. (1973): "Job Market Signalling", *Quarterly Journal of Economics* vol. 87, 355-79.

Starrett D. (1976): "Social Institutions, Imperfect Information, and the Distribution of Income", *Quarterly Journal of Economics* vol. 90, 261-84.

Stiglitz J. (1975): "The Theory of Screening, Education, and the Distribution of Income", *American Economic Review* vol. 65, 283-300.

Stiglitz J. (1987): "The Causes and Consequences of the Dependence of Quality on Prices", *Journal of Economic Literature* vol. 25, 1-47.

Taubman P. and T. Wales (1974): *Higher Education and Earnings*, New-York, McGraw Hill.

Tirole J. (1988): *The Theory of Industrial Organization*, Cambridge MA., M.I.T. Press.

U.S. Department of Labor Statistics (1968): *Employment and Earnings,* January Issue, Washington D.C.

Van Praag B., A Hagenaars, and W. Van Eck(1983): "The Influence of Classification and Observation Errors on the Measurement of Income Inequality", *Econometrica* vol. 51, 1093-1108.

Vickrey W. S. (1945): "Measuring Marginal Utility by Reactions to Risk", *Econometrica* vol. 13, 319-333.

Watts H. (1968): "An Economic Definition of Poverty", in D. Moynihan (ed.): *On Understanding Poverty*, New York, Basic Books.

Weiss A. (1991): *Efficiency Wages*, Oxford, Clarendon Press.

Weiss Y. (1985): "The Determination of Life Cycle Earnings: a Survey" in Ashenfelter O. and R. Layard (eds.): *A Handbook of Labor Economics*, Amsterdam, North Holland.

Whitmore G. (1978): "Statistical Tests for Stochastic Dominance", in Whitmore G. and M. Findlay (eds.): *Stochastic Dominance: an Approach to Decision-Making Under Uncertainty*, Toronto, Lexington Books.

Zimmerman D. (1992): "Regression Toward Mediocrity in Economic Stature", *American Economic Review* vol. 82, 409-429.