Chapter 1

The measurement of welfare and well-being, the Leyden approach

1.1 From economics to psychology

Modern economics is a discipline with numerous subfields, but nearly all relevant problems have to do with people and people's choice behavior. Individuals have limited resources and opportunities and therefore must choose between alternatives, for example, 1, 2, 3, ..., i, ... An efficient way to describe the choice problem is to attach a *utility value* to these alternatives, for example, \( U_1, U_2, U_3, \ldots, U_i, \ldots \) and to postulate that an individual chooses the alternative which has the highest utility value for him. For example, if there is a choice set \( \{1, 2, 3, \ldots, i, \ldots\} \), then the choice behavior is described mathematically by

\[
\max_{i=1,2,3,\ldots} U_i
\]

The implication of this description is that we can predict the individual's choice behavior if we know his utility values \( U_1, U_2, U_3, \ldots \) In empirical reality it is the other way round. We do not know the values of \( U \), but we may observe the choice process. If an individual consistently chooses alternative 1, we can infer that \( U_1 \) is larger than \( U_2, U_3, \ldots \). If we then remove alternative 1 from the choice set and 2 is chosen consistently we know that \( U_1 > U_2 \) and that \( U_2 \) is larger than other \( U \) values. In this way it is possible to find the preference ordering of the alternatives and to establish inequality relations between the \( U \) values. However, we are unable to say
whether $U_2$ is a little smaller than $U_1$ or whether $U_2$ is much smaller than $U_1$. In short, by observing choices we get an ordinal utility ordering.

The choice model may be extended in two ways. First, we can consider a set of alternatives that is a continuum. Alternatives can be described by a continuous variable $x$ or by more than one continuous variable, for example, $(x_1, x_2, \ldots, x_n) = x$. Then the utility values are denoted by the ordinal utility function $U(x)$. Secondly, we may assume that each decision maker $z$ has his own utility ordering. In that case, the ordinal utility function reads $U(x; z)$ where $z$ may incorporate individually varying parameters such as age, gender, income, social class, etc. We notice that this ordinal function is of the decision-utility type in the terminology of Kahneman, Wakker and Sarin (1997). It is needed to make decisions and is empirically established by observations of choice decisions.

The traditional example of choice behavior in economics is the purchasing behavior of consumers. The model starts from a utility function

$$U(x; z)$$

where $x$ stands for a vector of quantities of commodities purchased and $z$ for characteristics describing the individual's circumstances (e.g. age, gender). The consumer is faced by prices $p_1, \ldots, p_n$ for goods $x_1$ to $x_n$. If he has income $y$, his choice set is described by:

$$y \geq p_1 x_1 + \ldots + p_n x_n$$

Any commodity bundle $(x_1, x_2, \ldots, x_n)$ violating the constraint is too expensive for him. The behavioral model explains behavior by assuming that individuals maximize $U(x; z)$ with respect to the feasible commodity bundle $x$ subject to the freedom given by the choice set.

Edgeworth (1881) called $U(.)$ the utility function; Pareto (1904) called it the ophelimity function. Edgeworth more or less implicitly assumed that $U$ could be measured in a direct way. Samuelson (1945) therefore stated that "Edgeworth considered utility to be as real as his morning jam." Edgeworth interpreted $U$ as experienced utility, that is, a cardinal measure of the joy which the individual derives from the commodity bundle. Pareto became aware of the fact that it is difficult to establish the individual's utility function over goods. For the description of the consumer choice process, an experienced utility function appeared to be unnecessary. Actually, it is a choice between alternatives which may be described by

---

1Decision utility in the sense of Kahneman et al. (1997) may be ordinal or cardinal.
an ordinal utility function as described above. If $U(.)$ is an ordinal utility function, any other utility function which assigns the same ordering of utility to the alternatives is also a utility function describing that same choice process. For example, if $U_1 > U_2 > U_3 > 0$ describes the choice process between alternatives 1,2,3 then $\tilde{U}_1 = \sqrt{U_1} > \tilde{U}_2 = \sqrt{U_2} > \tilde{U}_3 = \sqrt{U_3}$ will describe the same process. Hence there is a whole equivalence class of ordinal utility functions describing the same preference structure.

It is an error to assume that Pareto denied the existence of cardinal utility or the possibility of measuring it, but he pointed to the fact that utility in the cardinal sense could not be measured by observing consumer behavior, and moreover that it was unnecessary to do so for consumer studies.

Robbins (1932), who had a tremendous influence on economics, was the first to proclaim that utility was immeasurable and that it was more or less a scientific folly to endeavor to measure it. At the very least, it should be left to psychologists.

Other economists such as Pigou (1948), and the Nobel laureates Tinbergen (1991) and Frisch (1932) were certainly of a different opinion.

However, the ordinal line has been continued by Arrow (1951) and Debreu (1959) who were able to include decisions over time and/or under uncertainty in this ordinal framework. They assume a preference ranking described by a utility function on the dated commodity space. Behavior is subject to a budget constraint where the consumption of goods and the prices of those goods are differentiated according to the date of consumption.

Similarly, they incorporate uncertainty by distinguishing states of nature $s$ varying over $S$ and commodities available only if $s$ prevails. Commodities are then available contingent on the status of nature, which is a priori not known to an individual. It can be shown that the model describes consumer choice behavior, but it is also clear that this model leads to a decision problem with an unworkable number of dimensions. Its realism as a positive behavioral model is not significant and it has never been used according to our knowledge in empirical work, except in very simplified versions.

In practice, economists are frequently confronted by problems where more is needed than the ordinal concept (see also Ng (1997)).

We think of decisions under uncertainty, the basis of insurance theory, investment and saving behavior. Decisions which have to do with different time periods such as saving and investment decisions need more than the ordinal concept. The objective function in such models is usually simplified to an additive form such as $\sum t w_t U_t$ or $\sum p_s U_s$ where the $U_t$ stands
for instantaneous period utilities and \( w_t \) for time-discounting weights, and where \( U_s \) stands for state-contingent utility and \( p_s \) for the (real or perceived) probability that state \( s \) occurs. Evidently, time-state mixtures and continuous generalizations are easy to think of.

There are two points of interest in these objective functions. The basic ingredient is a utility function \( U \) which is no longer ordinal. We cannot change the individual form at will according to a monotonous transformation. More specifically, maximizing \( \sum_t w_t \varphi(U_t) \) will yield an optimum which varies with \( \varphi(.) \), except if \( \varphi \) is a positive linear transformation (i.e., \( \varphi(.) = \alpha U + \beta \) with \( \alpha > 0 \)). The utility concept in this kind of problems is what economists call a cardinal utility function. It is a much smaller class which only allows for positive linear transformations.

Most mainstream economists have an uneasy feeling about cardinal utility functions. This uneasiness seems to be based on the Anglo-American attitude against cardinality instilled by Robbins. However, most actual studies conducted by economists start with very general ‘ordinal formulations’, but after a while they present a structural specification which in nine times out of ten turns out to be of the cardinal type (see also Van Praag (1968)). These cardinal utility functions are still of the ‘decision-utility’ type. They are instrumental to the description of decision processes.

There is a second class of problems for which economists need cardinal utility functions: normative problems. The first example of such problems arises if we try to look for optimal (re)distributions. Notably in income taxation, a progressive tax schedule (richer individuals pay relatively more tax than poorer individuals) is advocated so that the rich man suffers as much as the poor man. Such comparisons are impossible without a cardinal and interpersonally comparable utility function. Obviously, these utility functions are of the experienced utility type.

A second example is provided by equity measures: the concepts of a just income distribution, poverty and the evaluation of income inequality. It is evident that nearly all of these measures are based on a cardinal concept of experienced income utility, though this is rarely mentioned explicitly (c.f. Atkinson (1970)).

A third field where interpersonally comparable and cardinal utility is needed concerns all types of cost-benefit analyses, where specific measures such as building a bridge, deregulation of markets, specific health insurance programs, noise pollution by an airport, etc., have to be evaluated. In these cases, some citizens will profit and others will lose. Those benefits and costs may be partially translated in monetary amounts, but money means different things for different people. For example, when a policy means a loss of $100 to a poor man and a gain of $10,000 to a rich man,
it is not at all evident that the policy should be realized. The only way to make a decision is to create a balance in terms of comparable utility gains and losses.

The situation in economics is succinctly and wittily summarised by Wansbeek and Kapteyn.

"Utility seems to be to economists what the Lord is to theologians. Economists talk about utility all the time, but do not seem to have hope of ever observing it this side of heaven. In micro-economic theory, almost every model is built on utility functions of some kind. In empirical work little attempt is made to measure this all-pervasive concept. The concept is considered to be so esoteric as to defy direct measurement by mortals. Still, in a different role, viz., of non-economists, the same mortals are the sole possessors of utility functions and can do incredible things with it." (Kyklos, 36, pp. 249-269, 1983).

By detaching economics from the psychology of ‘feelings’, economists have found it difficult to have anything relevant to say on a whole range of issues. In the second part of this paper, we will review an attempt made by economists to measure utility functions using the evaluations given by individuals themselves. Before we do so however, we will first discuss the approaches taken in general to utility functions in the economic literature. We will divide the approaches that have been taken concerning the problem of utility functions in five distinct approaches.

1.2 General approaches to cardinal utility taken by economists

The first approach to cardinal utility, which is by far the most popular in the economic profession, is not to measure utility at all, but simply to assume a functional form of the utility function for the theoretical or empirical problem at hand. We will ignore this approach in the remainder of this chapter.

Economists who use the second approach, of which perhaps the best-known are Christensen et al. (1985) and Jorgenson and Slesnick (1984), have taken an axiomatic approach to utility functions. They specify the conditions they believe a utility function should satisfy and then derive (a
shape of) the utility function which fits these requirements. They then infer the level of utility that individuals enjoy from their observed behavior, which they then use to make normative statements. As utility levels are not directly measured in this approach, but essentially assumed, this approach is not elaborated here. Moreover, if this method has validity, it yields a cardinal decision utility.

Economists who take the third approach use subjective and objective indicators of the work and living conditions of individuals to define a measure of utility. This large group is subdivided into three groups: one group is concerned with the extent of poverty, another with the quality of life of nations, another with the quality of life of individuals.

The empirical literature on poverty centers around the material resources available to individuals (Townsend (1979, 1993), Sen (1987), Ravallion (1994)). The standard approach is to define households to be poor if their household income falls below a certain cut-off point. This cut-off point can be defined in several ways. For instance, in the ‘basic needs’ approach, the cut-off point is calculated from the expenditures needed to buy a basket of commodities that the researcher considers vital for individuals. In the relative needs approach, the cut-off point is defined to be a certain percentage of the average or median income in a country. It is clear that neither approach, which together form the bulk of the poverty literature, actually measures utility functions, but that they are based on the assumption that the utility of individuals whose income is below the cut-off point is in some sense ‘low’. See Callan and Nolan (1991) and chapter two for a more detailed review of the normative issues involved in poverty measurement.

Another branch of this literature examines the ‘quality of life’ of nations. In this literature (Kurian (1984), Nussbaum and Sen (1992), Sen (1987), Maasoumi (1989)), economists attempt to rank countries with respect to the quality of life. The quality of life is usually defined as a weighted average of specific country statistics. The statistics used include, for instance, the literacy level of the entire population, the literacy level of women, infant mortality rates, income levels per head, life expectancies of

---

2 See also Van Praag (1968) for an attempt to find a functional form of the utility function with the use of axioms and secondary assumptions.

3 The "quality of life" concept is very broad and interpreted by some to mean the same thing as happiness (e.g. Veenhoven (1996)), or average satisfaction (e.g. Dow and Juster (1985)). We discuss here the interpretation we believe most economists in this field use.
men and women, indicators of political stability, energy consumption per capita, average household size, the number of persons per physician, levels of civil liberties, etc. It is clear that these variables may be very important for the utility levels of individuals and nations, however the utility levels themselves are not measured by these variables. An obvious problem is then how one should weigh these statistics: does the quality of life increase more when the female literacy level increases by 1% or when the civil liberty index improves by 1%? It will be clear that if one does not want to use the evaluations of individuals themselves as a weighting method, the opinions of the researcher become the deciding criterion. The problem of how to weigh these different variables into a composite quality of life index is, not surprisingly, the main source of dispute in this literature. For an empirical analysis of some of the weighting methods employed, see Hirschberg et al. (1991).

Some of the works of Clark and Oswald (e.g. Clark and Oswald (1994)) belong to the last sub-group of the third category. In their 1994 paper, Clark and Oswald define 'unhappiness' by aggregating the answers to the following 12 questions:

1. Have you been able to concentrate on whatever you are doing?
2. Have you lost much sleep over worry?
3. Have you felt that you are playing a useful part in things?
4. Have you felt capable of making decisions about things?
5. Have you felt constantly under strain?
6. Have you felt you couldn’t overcome your difficulties?
7. Have you been able to enjoy your normal day-to-day activities?
8. Have you been able to face your problems?
9. Have you been feeling unhappy and depressed?
10. Have you been losing confidence in yourself?
11. Have you been thinking of yourself as a worthless person?
12. Have you been feeling reasonably happy, all things considered?

The variable ‘unhappiness’ ranges from 0 to 12, whereby 12 denotes the maximum level of unhappiness and 0 a complete lack of unhappiness. Although some of these questions could arguably be seen as a measure of utility, such as questions 9 and 12, the simple aggregate of all 12 questions cannot be seen as a direct measurement of utility: utility is an evaluation of an individual of his circumstances. Although ‘losing a lot of sleep’ or ‘being under strain’ may affect utility or may be affected by utility, they do not directly measure a utility level for they are not an evaluation of ‘losing sleep’ or ‘being under strain’. This measure of happiness may cor-
relate perfectly with the experienced utility of individuals and may hence be as useful as any other measure of experienced utility. Nevertheless, it remains an indirect measure of experienced utility which is useful only if ‘losing sleep’ and ‘being under strain’ correlate with experienced utility (which seems very likely). Hence, it is a measure of the quality of life entirely on its own. Clark and Oswald seem to acknowledge this by arguing that the individual scores are ‘more accurately’ described as ‘mental stress’ scores. Other individual measures of the quality of life of an individual which are based on aggregations of individual circumstances also fall into this category.

A fourth approach is to estimate decision utility functions by performing probability-choice experiments on individuals. When individuals must choose between either a certain outcome $Y$ or a lottery in which fate decides whether they will receive an outcome less than $Y$ or an outcome greater than $Y$, individuals will reveal the relative attractiveness of the sure $Y$ versus the proposed lottery. The main problem in this line of research has been that individuals are not good at using probabilities: they overestimate small probabilities and underestimate large probabilities, as was first demonstrated by the Allais paradox (see Allais and Hagen (1979)). This means that the choice of an individual for a lottery is the result of a combination of the individual’s valuation of the outcomes and of the individual’s perception of the probability of the outcomes. Following the theoretical advances by Kahneman and Tversky (1979) and Wakker and Tversky (1993), Kahneman et al. (1991) have managed to isolate the effect of gains and losses on individual’s evaluation of outcomes. We will ignore the results on probabilities and focus on the value function they find. The shape of the value function suggested by the choice experiments of Kahneman et al. (1991) is sketched in figure 1.

The main feature of this function is that losses are found to have a greater impact than gains. A second characteristic of this value function is that it levels off at either end of the loss or gain scale, which implies decreasing marginal value of losses and gains. It implies a convex-concave shape, also suggested by Markowitz (1952) and Van Praag (1968).

Finally, the fifth approach taken is to assume that individuals are able to describe their utility level by means of verbal qualifications. The rest of this chapter will be devoted to the attempts of economists who belong to the Leyden group where this approach was initiated. There are of course also other economists who use verbal qualifiers as measures of experienced utility (e.g. Clark (1996), Clark and Oswald (1996), Heywood et al. (1997),
1.3 Utility measurement based on verbal qualifiers

1.3.1 The Leyden approach

In this section we will discuss an economic line of research which tries to operationalize the concept of experienced utility. It originated at Leyden University in the Netherlands in the early 1970s. Its main contributors are Van Praag, Kapteyn, Wansbeek, Hagenaars, Van der Sar, Plug, and Frijters. It is known in the literature as the Leyden approach (or school). For psychologists, the ideas in this approach may not appear alien, but for most economists they were and still are. Most economists still believe that cardinal experienced utility is unmeasurable and that any measurement should be based on observed decision behavior. This meant that the Leyden approach met with stiff opposition, disbelief and outright hostility. The most outspoken example of this attitude is found in an article by Seidl (1994) in the European Economic Review where he criticises Van Praag (1968).^4

Although Van Praag (1968) served as a theoretical basis, the ensuing literature on the Leyden approach started with Van Praag (1971) and is

---

^4A reply was given by Van Praag and Kapteyn (1995).
mainly empirical and data-oriented. The Leyden approach focuses primarily on the evaluation of income, although in later work the focus was extended. We also speak of utility of income, income satisfaction, or, in other words, economic welfare. We drop the adjective economic from now on, but when we use the term welfare, we have welfare derived from income in mind. This concept is narrower than the concept of well-being which includes feelings associated with factors unconnected to income or purchasing power. In Section 7 we shall consider well-being and its relationship with welfare in greater detail.

The Leyden approach is based on two assumptions. The first is that individuals are able to evaluate income levels in general and their own income in particular in terms of ‘good’, ‘bad’, ‘sufficient’, etc. We call these terms verbal qualifiers. The second assumption is that verbal labels can be translated in a meaningful way into a numerical evaluation on a bounded scale, for example [0,1]. We shall consider both steps of the measurement procedure in detail.

If we are interested in how a specific income level is evaluated, there are two ways to gather information. The first and most natural way is to propose a sequence of income levels and to ask for their verbal qualifications. An example of this type of question follows:

"Here is a list of income levels per month, after tax: please evaluate these amounts using verbal qualifications, such as ‘very bad’, ‘bad’, ‘insufficient’, ‘sufficient’, ‘good’, ‘very good’:

$2000
$4000
$6000
$8000
$10,000"

It is obvious that someone who earns $20,000 a month would be unable to make a distinction between most of these levels. All the incomes are insufficient or worse for him. Therefore, instead of starting with income levels, we can also supply the verbal qualifications as stimuli and ask the individual respondent which income level corresponds with the verbal label. This leads to the so-called Income Evaluation Question:

"while keeping prices constant, what after-tax total monthly income would you consider for your family as:"
The measurement of welfare and well-being, the Leyden approach

very bad,........................................... $\ldots$
bad, .............................................. $\ldots$
sufficient,............................................... $\ldots$
insufficient,......................................... $\ldots$
good, ............................................... $\ldots$
very good, .......................................... $\ldots$

This question appears to have been successful in anonymous mail questionnaires, although it has also been posed orally with success. Theoretically, finding a continuous relationship between income and utility would require an infinite number of levels, but, in practice, between four to nine levels have been and can be used. We will discuss here the format used most often, the six-level format.

The question is now how we derive a welfare function from the answers to this question. Or more precisely, how we translate the verbal labels into numbers on a [0,1] scale. Following Van Praag (1971), we make an assumption about the way individuals fill out the question. We assume that respondents try to provide information to the interviewer about the shape of their welfare function. The most accurate way for individuals to provide information then depends on the accuracy criterion. Van Praag (1971) and Kapteyn (1977) show that, for a broad class of intuitively plausible criterion functions, the best way for a respondent to provide information is to choose the answers in such a way that each of the six levels corresponds to a jump of 1/6. This is the so-called equal quantile assumption (EQA). It implies that

\[
\begin{align*}
U(\text{very bad}) & = U(\text{first interval}) = 1/12 \\
U(\text{bad}) & = U(\text{second interval}) = 3/12 \\
& \quad \vdots \\
U(\text{very good}) & = U(\text{last interval}) = 11/12
\end{align*}
\]

It may be surmised that, even if the verbal descriptions are somewhat vague, the respondent will tend to interpret the question as if it were an equal partition. Only if the verbal labels are ambiguous, practically equal
or strongly suggest an unequal partition, we should no longer expect this effect.

If the number of verbal labels is \( k \), the general formula for the welfare corresponding to the \( j^{th} \) verbal label is obviously \( \frac{2k-i}{2k} \). This reasoning and the EQA assumption are very similar to the thesis developed by Parducci (see e.g. Parducci (1995)). It is obvious that this translation of verbal labels into numbers is a linking pin in this measurement procedure. It has been subject to criticism by some economists, while on the other hand experimental psychologists do not find much to criticize: it is a type of Thurstonian measurement. If we do not accept this or any translation into figures, it is obvious that a meaningful analysis of the response is severely hampered, although not impossible (see later).

In Van Praag (1991), an experiment is described in which five labels were supplied and 364 respondents were asked to 'translate' these verbal labels onto a \([0,100]\)-scale. Similarly, the same labels had to be linked with line segments. Both the numbers between \([0,100]\) and the lengths of the line segments were re-scaled onto a \([0,1]\) mapping. We present the average results for 364 respondents in Table 1.

### Table 1: Translation into numbers and line segments

<table>
<thead>
<tr>
<th>Numbers:</th>
<th>Empirical mean</th>
<th>St. dev.</th>
<th>Theor. pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td>very bad</td>
<td>( \bar{\nu}_1 = 0.0892 )</td>
<td>0.0927</td>
<td>0.1</td>
</tr>
<tr>
<td>bad</td>
<td>( \bar{\nu}_2 = 0.2013 )</td>
<td>0.1234</td>
<td>0.3</td>
</tr>
<tr>
<td>not bad, not good</td>
<td>( \bar{\nu}_3 = 0.4719 )</td>
<td>0.1117</td>
<td>0.5</td>
</tr>
<tr>
<td>good</td>
<td>( \bar{\nu}_4 = 0.6682 )</td>
<td>0.1169</td>
<td>0.7</td>
</tr>
<tr>
<td>very good</td>
<td>( \bar{\nu}_5 = 0.8655 )</td>
<td>0.0941</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Line segments</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>very bad</td>
<td>( \bar{\omega}_1 = 0.0734 )</td>
<td>0.0556</td>
<td>0.1</td>
</tr>
<tr>
<td>bad</td>
<td>( \bar{\omega}_2 = 0.1799 )</td>
<td>0.0934</td>
<td>0.3</td>
</tr>
<tr>
<td>not good, not bad</td>
<td>( \bar{\omega}_3 = 0.4008 )</td>
<td>0.1056</td>
<td>0.5</td>
</tr>
<tr>
<td>good</td>
<td>( \bar{\omega}_4 = 0.5980 )</td>
<td>0.1158</td>
<td>0.7</td>
</tr>
<tr>
<td>very good</td>
<td>( \bar{\omega}_5 = 0.8230 )</td>
<td>0.1195</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Source: Van Praag (1991)

For the ‘numbers’ case, one can see that all averages fall within a one \( \sigma \)-interval of their theoretical prediction. This also holds for all levels for the line segments, except one. It is intriguing that the averages are all below their theoretical prediction. Perhaps this is due to the order in which the verbal labels were supplied. We think, but do not know, that the bias would have been the other way around if the order in which the verbal labels were supplied was reversed. When we regress the translation of the verbal labels into numbers by individual \( i \), say \( \nu_{i,n} \), onto the translation
of the verbal label into a line-segment, say \( w_{i,n} \), we find (st. deviations in brackets)

\[
\nu_{i,n} = 0.056 + 0.974w_{i,n} \\
(0.005) (0.010)
\]

\[
R^2 = 0.848
\]

for 364*5 observations, where we did not account for the fact that the five level disturbances per individual will be strongly correlated. The fit is however remarkably good. From Table 1 and this regression, we can draw some tentative conclusions:

1. A verbal label sequence seems to be understood in a similar way by different respondents, irrespective of the context of the individual respondent.

2. A verbal label sequence may be translated on a numerical scale or on a line scale: in both cases the translations are uniform over individuals.

3. Translations via various translation mechanisms (lines and figures) are consistent with each other. That is, we seem to be measuring the same thing, irrespective of whether we use line segments or numbers.

4. The verbal labels are translated on a bounded scale roughly in accordance with the Equal Quantile Assumption.

An interesting point is that these results were found in a context-free setting, that is, the respondents did not know which concept they were evaluating.

A final point of critique is whether the verbal labels ‘good’, ‘bad’, etc., convey the same feeling to every respondent. If not, we falsely assume that individuals derive the same degree of joy from their income, when describing the same verbal label. Actually, this is a question of psycho-linguistics. Generally, the basic idea of language is that frequently used words will have the same meaning and emotional connotation for the members of a language community. It is the main tool of communication between people. Hence, we must assume that verbal labels like ‘good’, ‘bad’, etc., mean approximately the same thing to all respondents sharing the same language. In chapter 7, this issue will be reviewed more thoroughly.
1.3.2 The shape of the welfare function

For each respondent we now have six income levels connected to six utility levels. The shape of the function can be inferred from these six combinations. Many functions can be fitted using these six points. In Van Praag (1968), it was argued on theoretical grounds that it would be a lognormal distribution function. The reason to use a distribution function is that we assume boundedness of the utility function: there is a worst and a best position in terms of welfare (satisfaction). It is also known that the Von Neumann-Morgenstern model requires a bounded utility function (see Savage (1954)).

Van Herwaarden and Kapteyn (1981) showed that the points of the welfare function, which were found empirically, best fitted a lognormal curve within the class of distribution functions. The logarithmic function did slightly better, but it is not bounded. Also, the logarithmic function is not borne out by the choice experiments of Kahneman et al. (1991) and others: the marginal effect of greater losses is found to decrease, whereas the logarithmic function would imply that they should increase.

The lognormal function is defined as

\[ \Lambda(y; \mu, \sigma) = N(\ln y; \mu, \sigma) = N\left( \frac{\ln y - \mu}{\sigma}; 0, 1 \right) \]

where \( N(\cdot; 0, 1) \) stands for the standard lognormal distribution function. The lognormal function is sketched in Figure 2. Notice the resemblance to the shape suggested by the experiments of Kahneman et al. (1991): in both cases the function is S-shaped. Also, it is generally the case that losses to an individual have a greater effect than gains.\(^5\)

The parameter \( \mu \) is interpreted by realizing that \( \Lambda(e^{\mu}; \mu, \sigma) = 0.5 \). Hence, the income level \( e^{\mu} \) is halfway between the worst and the best situation.

There are two interesting aspects about this function. First, the function is not concave for all income levels, but convex for low incomes. This runs counter to mainstream economic assumptions. In economics, it is conventional wisdom that the utility function of income is always concave. This is known as the so-called Law of Decreasing Marginal Utility, also known as Gossen’s first law. It has always been based on introspection.

\(^5\)One particular feature of the value function found by Kahneman et al. (1991) cannot be replicated: Kahneman et al. find a value function which changes direction abruptly at the reference position. The number of levels used in our measure is simply too small to find such a jump in direction.
Concavity implies that individuals are risk-averse, but scientific experiments with insurance or gambling behavior show that this is not always true; it therefore follows that a utility function may be convex in certain regions.\(^6\)

The second point of interest about the lognormal utility function consists of the two parameters \(\mu\) and \(\sigma\) which may vary individually. Two functions with different \(\mu\) and equal \(\sigma\) are sketched in Figure 3a. In Figure 3b two functions are sketched with different \(\sigma\) and equal \(\mu\).

One can see that as \(\mu\) increases, the individual needs more income to reach the same welfare level. For instance, in order to reach the welfare level 0.5, the person A with \(\mu_A = \ln(4000)\) needs $4000 per month, while B needs $6000 per month to reach the same welfare level. If the welfare levels of individuals A and B are to be equal for other welfare levels (if \(\sigma\) is equal for both persons), it should hold that

\[
\ln y_A - \mu_A = \ln y_B - \mu_B
\]

Hence, for any welfare level, income levels are equivalent to A and B if

\[
\ln \frac{y_A}{y_B} = \mu_A - \mu_B
\]

\(^6\)A variable of much economic interest, Pratt’s (1964) measure of relative risk aversion (or Frisch flexibility) can be directly calculated as

\[
\frac{\partial \ln u}{\partial \ln y} = -\frac{1}{2\sigma^2}(\ln y - \mu) - 1
\]

It varies from highly positive for small \(y\) to very negative for large \(y\).
and therefore

$$\frac{y_A}{y_B} = e^{\mu_A - \mu_B} = \frac{4000}{6000}$$

Hence, a change in $\mu$ implies a proportional shift of the welfare function. One of our main preoccupations in the remaining section is to discover why individual's $\mu-$values differ.

The parameter $\sigma$ defines the slope of the welfare function.
In Figure 3b, two functions are sketched with \( \sigma_A < \sigma_B \). If \( \sigma = 0 \) we get the limiting case where individuals are completely unsatisfied with any income until their income reaches \( e^\mu \), and where they are completely satisfied if income exceeds \( e^\mu \). It is the welfare function of a Hermit. The parameter \( \sigma \) is called the welfare sensitivity of the individual.

The parameters \( \mu \) and \( \sigma \) are estimated for each individual by

\[
\hat{\mu} = \frac{1}{6} \sum \ln c_j \quad \text{and} \quad \hat{\sigma} = \frac{1}{5} \sum (\ln c_j - \mu)^2
\]

where \( c_1, \ldots, c_6 \) stand for the six income levels reported in the IEQ.

1.3.3 The definition of income

In the usual IEQ version, the income concept is after-tax monthly household income. In some versions income per year has also been used and/or before tax income (see Dubnoff et al. (1981)). The choice of the definition should be adapted to what is well-known to the individual. Hence, an entrepreneur who knows his annual income better than his monthly income should be questioned in terms of his annual income, while a civil servant who is paid monthly should be approached in terms of his monthly income.

1.4 The explanation of the welfare function

In mainstream literature, it is always assumed that the utility function of income is the same for all individuals. A major finding of our empirical research, although intuitively completely plausible, is that individual welfare functions differ between individuals. When differences are found, the imminent question is whether such differences are structural and can be correlated with observable variables. In our case, this means that we try to 'explain' the variable \( \mu \) by other factors, varying per individual and/or environment. In the studies, it appeared that \( \mu \) could be explained to a large extent\(^7\). The parameter \( \sigma \) posed much more of a problem. We shall therefore concentrate on the explanation of \( \mu \) and assume that \( \sigma \) is constant.

We recall that \( \mu \) determines the position of \( U(y) \). If \( \mu \) increases, the individual becomes less satisfied with the same amount of income. In other words \( U(y;\mu) \) is decreasing in \( \mu \). The first determinant that naturally comes to mind is the size of the family to be supported from the income. Income needs are probably also determined by the actual circumstances of

\(^7\)An explanation does not necessarily mean a one-way causal relationship.
the individual, for instance as reflected by the individual’s current income \( y_c \). We therefore expect that needs will increase with family size (denoted by \( fs \)) and with current income \( y_c \). Hence, \( fs \) and \( y_c \) are parameters in the individual welfare function. In Van Praag (1971) and Van Praag and Kapteyn (1973), the following simple relation has been found

\[ \mu_i = \text{constant} + \beta_1 \ln fs_i + \beta_2 \ln y_{i,c} \]  

(1.4.1)

In Van Praag and Kapteyn (1973), the following (approximate values) were found \( \beta_1 = 0.1 \) and \( \beta_2 = 0.6 \), \( R^2 = 0.6 \), where \( fs_i \) denotes the number of individuals living in the household of respondent \( i \), and \( y_{i,c} \) denotes the current household income of \( i \).

Since then, the IEQ has been posed in many countries, and similar results have been found. We give an example drawn from a study on poverty by Van Praag, Hagenaars and Van Weeren (1982), based on a 1979 EUROSTAT survey of eight European countries. Moreover, we add values for Russia estimated by Frijters and Van Praag (1995, see also chapter 2). In Table 2, we present the regression estimates for the nine countries using the equation

\[ \mu_i = \beta_0 + \beta_1 \ln fs_i + \beta_2 \ln y_{i,c} + f(X_i) + u_i \]

where \( X \) denotes a number of variables used in the regression which we do not show (including age, education, employment levels and gender), and \( u_i \) denotes the normally distributed error term.

<table>
<thead>
<tr>
<th>Countries</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( N )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>0.097</td>
<td>0.433</td>
<td>1272</td>
<td>0.695</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.075</td>
<td>0.631</td>
<td>1972</td>
<td>0.829</td>
</tr>
<tr>
<td>France</td>
<td>0.069</td>
<td>0.505</td>
<td>2052</td>
<td>0.676</td>
</tr>
<tr>
<td>W. Germany</td>
<td>0.112</td>
<td>0.583</td>
<td>1574</td>
<td>0.693</td>
</tr>
<tr>
<td>Great Br.</td>
<td>0.115</td>
<td>0.364</td>
<td>1183</td>
<td>0.575</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.169</td>
<td>0.455</td>
<td>1733</td>
<td>0.636</td>
</tr>
<tr>
<td>Italy</td>
<td>0.156</td>
<td>0.381</td>
<td>1911</td>
<td>0.510</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>0.100</td>
<td>0.537</td>
<td>1933</td>
<td>0.664</td>
</tr>
<tr>
<td>Russia 1995</td>
<td>0.250</td>
<td>0.501</td>
<td>1444</td>
<td>0.501</td>
</tr>
</tbody>
</table>

Source: Van Praag, Hagenaars and Van Weeren (1982) + own calculations

All coefficients are highly significant. The variables vary over the nine countries but not dramatically. The value of \( \beta_1 \) is related to the costs of children to the household and will will
depends on the level of state provisions. If childcare, education and other services used by children are free, then we may expect a $\beta_1$ of about zero. On the other hand, in poor countries with a less liberal system, $\beta_1$ may be rather high. This is indeed what we observe: the highest coefficient of $\beta_1$ is for Russia in 1995 where state support for families with children is virtually non-existent.

It is not surprising that the satisfaction derived from a specific income level depends on the size of the household. Somewhat more surprising, especially for most economists, is that income satisfaction for any income level, not only for an individual’s own current income, depends on an individual’s own current income. It implies that two individual A and B with current incomes $y_{A,c}$ and $y_{B,c}$ will evaluate any income differently. More precisely, the following is usually true:

$$U(y_{B};fs,y_{B}) \neq U(y_{B};fs,y_{A})$$

That is, B evaluates his own income differently than A would evaluate the income of B. It is obvious that this fact is very relevant for the evaluation of social inequality, for the theory of a fair income distribution and for the evaluation of social welfare. The outcomes of such normative evaluations depend on the income norm of the evaluations. Actually, $U_A(y;fs_A,y_{A,c})$ describes the norms of A with respect to what equals a 'bad', ..., 'good' income and all levels in between.

A person’s income may increase, for example from $y_1^{(1)}$ to $y_1^{(2)}$. The evaluation of this change will be evaluated differently before the change and after the change, or, as economists say, ex ante and ex post. The ex ante evaluation of future income is $U_A(y_1^{(2)};fs,y_1^{(1)})$, while the ex post evaluation is $U_A(y_2^{(2)};fs,y_2^{(1)})$. We sketch the difference between the ex ante and ex post welfare function in figure 4.

Due to the fact that $\mu$ increases with the income change, the welfare function shifts to the right. The effect of this is that the ex post evaluation of both $y_1^{(1)}$ and $y_1^{(2)}$ falls compared to the corresponding ex ante evaluations. It can be seen, and also shown, that the ex ante welfare gain is larger than the ex post gain. As a consequence, the ex ante evaluation is exaggerated when reconsidered later on, or to put it differently, the income increase will be a disappointment in retrospect. The value of the coefficient $\beta_2$ is crucial in this context. If $\beta_2 = 0$, the curve will not shift to the right and the whole income increase will be translated as a welfare increase. In that case ex ante and ex post evaluations are equal.

On the other hand, if $\beta_2 = 1$, perceived welfare will not increase at all.
This can be seen by examining
\[ \ln y_c - \mu = \ln y_c - \beta_0 - \beta_1 \ln f s - 1.00 \ln y_c = -\beta_0 - \beta_1 \ln f s \]

In this case, the subjective ex post welfare evaluation does not depend on actual income. Evidently, this is a pathological case which has not been found in reality. The anticipated welfare increase would end with a complete deception.

The phenomenon of a shifting welfare function arising from a partial adaptation of income norms to changing current incomes, is what Brickman and Cambell (1971) called the hedonic 'treadmill'. Van Praag (1971) introduced the term 'preference drift' for the same phenomenon.

If all individuals have their own norms with regard to income levels, which depend on their own circumstances, the question is justified whether it is possible to construct social standards with respect to what is a 'good' income, a 'bad' income, etc. This is possible in a certain sense. We define a social standard for a 'good' income, say \( y_{good} \), as that level of income which is evaluated to be 'good' by an individual with that current income. If a 'good' income corresponds with a welfare value of 0.7 on a [0,1] scale, it implies that \( y_{good} \) is the solution to the equation

\[ U(y_{good}; f s, y_{good}) = 0.7 \]

Using lognormality and our estimate of \( \mu \), it is possible to show that

\[ U(y_{good}; f s, \tilde{y}_{good}) = \Lambda(y_{good}; \frac{\beta_0 + \beta_1 \ln f s}{1 - \beta_2}, \frac{\sigma}{1 - \beta_2}) \]
Similarly, we can obtain a social standard income for each possible welfare level, sketched in the next figure.

![Figure 5: The social standard welfare function](image)

We call the ensuing welfare function of the social standard income levels, which is also lognormal, a *social standard function*. We know that someone with $y_{0.4}$ current income will evaluate his own income by 0.4. This analysis is frequently used to define a *subjective poverty* line as $y_{0.4}$ for poverty and $y_{0.5}$ as near-poverty. Notice that this line varies as a function of family size. Hence, there is a two-person household poverty line, a three-person poverty line, etc. The social standard function is an obvious tool for social policy and the evaluation of income redistribution and tax policy.

From a social-psychological viewpoint, it is very interesting to compare the welfare sensitivity of the *individual* welfare function $\sigma$ with the corresponding slope parameter of the social standard function $\frac{\sigma}{1-\beta_2}$. If $0<\beta_2<1$, the latter function is less steep than its individual counterpart. In other words, the larger the preference drift $\beta_2$, or in psychological terms, the stronger the working of the hedonic treadmill, the flatter the social standard curve will be compared to the individual welfare function.

Obviously, the difference between the two functions explains why a person with moderate income, for instance $40,000 a year, thinks that someone with $100,000 is rich while the rich person himself with $100,000 does not perceive himself to be rich. In the same way, people with $20,000 do not feel as poor as the observer earning $40,000 thinks they would.

The explanation of $\mu$ by individual variables and the stability of these explanations over samples (see also Van Praag and Van der Sar (1988))
may be seen as indirect evidence for the validity of the Welfare Function of Income. The measured concept may be explained to a certain extent by individual circumstances in a plausible way. One of the more recent additions is a quadratic part in age. It is seen that financial needs are greatest at the age of about 40.

However, the explanation of $\mu$ may be useful for policy purposes as well. If we find that the welfare derived from income depends on family size, this gives a natural clue to the question what family allowance would keep the family at the same household level if family size is increased from two to three by having a child: the welfare a household derives from income is

$$U_{ind} = \Lambda(y_{i,c}; \text{constant} + 0.1 \ln f + 0.6 \ln y_{i,c}; \sigma)$$

In order to keep welfare constant if $f$ increases from 2 to 3, we should add $0.1 \ln(\frac{3}{2})$ to $\ln y_{i,c}$ or multiply $y_{i,c}$ by $(\frac{3}{2})^{0.1}$.

However, in the long run, this increase will not be enough to compensate the family for an increase in family size, as current income $y_{i,c}$ increases and hence $\mu$. Therefore, we need a second increase of $0.1 \ln(\frac{3}{2}) \times 0.6$ and so forth. The total increase necessary to compensate the household equals

$$0.1 \ln(\frac{3}{2}) \left[1 + 0.6 + 0.6^2 + \ldots \right] = \frac{0.1 \ln(\frac{3}{2})}{1 - 0.6}$$

and this is precisely what the social standard welfare function would prescribe. Here we encounter a dynamic aspect, viz., that the individual welfare function is anchored on own current income. That is the meaning of preference drift. People adapt their norms to the present situation.

Parts of this analysis are also possible on the separate $c_i$ levels without any reference to a cardinal utility function (see Van Praag and Van der Sar (1988)).

### 1.5 Dynamics

In the previous section, we described how the need parameter $\mu$ could be explained by variables such as family size and current income $y_c$. The latter effect is now refined by supposing that $\mu$ depends not only on present income but also on income in the past and income that is anticipated in the future. It follows that in the $\mu -$equation we replace $y_c$ by $\ldots,y_{-2},y_{-1},y_0,\hat{y}_1,\hat{y}_2,\ldots$ whereby $y_0$ denotes current income, $y_{-2},y_{-1}$, for incomes one or two years in the past and where $\hat{y}_1$ stands for anticipated future income in one year’s time. All experienced and anticipated incomes contribute to
the formation of our present norm on incomes. In its simplest form, the $\mu$-equation looks like

$$\mu_i = \beta_0 + \beta_1 \ln f s_i + \beta_2 \left( \sum_{t=-\infty}^{+\infty} w_t \ln y_{i,t} \right)$$

where $i$ refers to respondent $i$.

The coefficients $w_0, w_1, w_2, \ldots$ are weights which add up to one, whereby the weight $w_0$ denotes the weight of the present income, while $w_p = \sum_{t=-\infty}^{-1} w_t$ and $w_f = \sum_{t=1}^{\infty} w_t$ denote the weight of all past incomes and anticipated future incomes respectively. Van Praag and Van Weeren (1983,1988) estimated the parameters of this model on Dutch panel data. The main question concerns how the distribution of time weights will look. They regressed $\mu_i$ on the incomes of the three years in which the panel was held. For the second wave they found

$$\mu_i = 3.04 + 0.10 \ln f s_i + 0.68(0.16 \ln y_{i,t-1} + 0.75 \ln y_{i,t} + 0.09 \ln y_{i,t+1})$$

$R^2 = 0.69$

$N = 645$

where all coefficients are significant. The results tell us that current income has the greatest time-weight, which implies that the time-weight distribution peaks near the present. Also, incomes in the past carry more time weight than incomes in the future, which suggests that on aggregate the time-weight distribution peaks just before the present. Of course, this is an aggregate relationship which will differ for individuals of different ages and education profiles. For a more complete analysis, more incomes than the three available were needed. Therefore, Van Praag and Van Weeren (1988) used econometric techniques to estimate the incomes which were further back than one year, i.e. $y_{-3}, \ldots, y_{-2}$. They also estimated incomes further than one year in the future, i.e. $y_2, y_3, \ldots$ With the use of this complete income stream, they looked somewhat further at the shape of the time-weight distribution. In general, they found the time-weight distribution to have the shape of a normal curve. More specifically, the time-weight distribution may be characterized by a mode parameter, $\mu_\tau$, and a dispersion parameter, $\sigma_\tau$. The empirically estimated shapes of the time weight distribution are presented in Figure 5.1 for three age brackets, viz., at 30, 50 and 70.

8The likely result of using estimates for some incomes is that the effect of income different from the present income will be underestimated. The qualitative results should, however, remain the same.
The most interesting points are that:

- The time weight distribution varies for different ages.
- The distribution is not symmetric around the present.
- The time weights of the past are greatest for young and old people.
- The middle age bracket derives its norm mostly from the present and the anticipated future.
- The dispersion of the distribution varies considerably over different ages. In midlife, the time weights are extremely concentrated.

The mode and symmetry point of the time weight distribution is at $\mu_\tau$. We call it the *time focus* of the individual. It shifts from more than one year in the past, (-1.3) at 20, to almost half a year in the future, (0.45) at 50, while it shifts back to the past, (-0.43) for the age of 70.

The change of $\sigma_\tau$ is also interesting. We call $\sigma_\tau$ the *time span* of the individual. It is rather long for young individuals and shortens when people approach midlife. The time span is intimately related to the *velocity of time* as it is perceived by the individual. The midlife has a narrow time horizon which implies that the individuals then live "by the day". The velocity of life is high. For young and old people, the time horizon is wider and hence the velocity of time is lower. We call the reciprocal of $\sigma_\tau$, i.e. $\frac{1}{\sigma_\tau}$ the *subjective velocity of time*. 

---

*Figure 6*
In Table 3, we present the relevant figures for several age classes. It is seen that the subjective velocity of time $\frac{1}{\sigma_T}$ increases by a factor $\frac{1.44}{0.09} \approx 15$ up to midlife, and then falls by a factor 6 at age 70, and still more at later ages.

<table>
<thead>
<tr>
<th>age</th>
<th>$\mu_T$</th>
<th>$\sigma_T$</th>
<th>$w_p$</th>
<th>$w_0$</th>
<th>$w_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>-1.32</td>
<td>1.44</td>
<td>0.72</td>
<td>0.18</td>
<td>0.10</td>
</tr>
<tr>
<td>30</td>
<td>-0.32</td>
<td>0.71</td>
<td>0.40</td>
<td>0.48</td>
<td>0.12</td>
</tr>
<tr>
<td>40</td>
<td>0.27</td>
<td>0.26</td>
<td>0.00</td>
<td>0.81</td>
<td>0.19</td>
</tr>
<tr>
<td>50</td>
<td>0.45</td>
<td>0.09</td>
<td>0.00</td>
<td>0.70</td>
<td>0.30</td>
</tr>
<tr>
<td>60</td>
<td>0.22</td>
<td>0.21</td>
<td>0.00</td>
<td>0.91</td>
<td>0.09</td>
</tr>
<tr>
<td>70</td>
<td>-0.43</td>
<td>0.62</td>
<td>0.46</td>
<td>0.48</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Source: Van Praag and Van Weeren (1988)

The time weight distribution is clearly important for individuals because it determines the speed of adaptation of the income norms of the individual when faced with changing circumstances. This may be the case for individuals who become jobless and then become dependent on social benefits. The adaptation process may be a reason to smooth the path of the income reduction over time, in order to smooth the decline in welfare.

Another potential application is to evaluate the impact of inflation and accelerating inflation on the income norms and the satisfaction level derived from income. These applications are discussed in greater detail in Van Praag and Van Weeren (1988).

The research on the time weight distribution has not been repeated since 1988. Therefore, this must be seen as a first attempt, where the results have to be considered with care. It may be that other models would yield other results. This method of obtaining time weights is based on a simple household survey and is very cheap compared to experimental laboratory experiments.

The estimates of time weights are exclusively based on the analysis of income norms. The memory and anticipation weights from norms on other subjects, for instance on fashion, housing, ethics, may be determined by other variables and have different time weight distributions. There is a great need for more research in this area and cooperation with psychologists.
1.6 Reference groups? The social filter

A major concept in sociology is the reference group. It is a matter of fact that opinions and norms are not formed by individuals in splendid isolation, but that their norms and opinions are influenced to a large extent by parents, peers and others. They form the social reference group of the individual. When we consider society as a two-dimensional space, where each person is characterised by age and income, we may think of the reference group of person A, say $G_A$, as a region in that space and similarly we may sketch the reference group of B. In figure 7 two such reference spaces are described by elipsoids.

![Figure 7: the social reference group](image)

In this figure, B is not in the reference group of A, but A is in the reference group of B. In practice it is difficult to draw exact borderlines. It may be safe to assume that not all individuals in $G_A$ carry equal weight for A. We may even assume that it is impossible to draw a borderline, but that in a sense the whole society influences the norms and opinions of A, although some people are "far out" and carry a weight near zero, while other individuals carry a heavy weight. This leads to a generalisation of the concept of reference group. We replace it by a reference weight distribution. The weight distribution may differ for each individual. For norms on food, clothing, visiting the theatre, one's religion, the intensity of religious observance, etc., we may expect different reference weight distributions.

In this context we concentrate on income norms. As we saw, the individuals income norm is explicitised by his response to the IEQ and described by his welfare function of income.
How are norms formed? One hypothesis is that most norms with respect to quantitative variables are *purely relative*. For income in particular it may be that people evaluate income levels by estimating what fraction of income earners earns less than themselves and what fraction earns more.

For instance, if somebody believes that 80% earns less he feels very well-off with his income, while if he assumes that only 20% earns less, he feels miserable. This hypothesis is termed the preference formation hypothesis by Kapteyn (1977). Layard (1980) used a similar hypothesis.

If everybody had the same income distribution in mind, we would have

$$U_i(y) = G(y)$$

where $U_i(.)$ is the individual welfare function of individual $i$ and $G(.)$ is the cumulative income distribution function of the population. In practice $U_i(.)$ differs between individuals, as we saw before. This implies that individuals do not refer to the same population, which means we interpret $U_i(y)$ as the income distribution function of the reference group, the so-called *perceived* income distribution. We sketch the density functions corresponding to both distributions in figure 8.

![Figure 8: Objective distribution density vs. perceived income distribution density.](image)

The objective distribution counts numbers, i.e., everybody carries equal weight. The perceived distribution gives some income classes (all incomes between A and B) a greater weight, whilst those individuals with incomes less than A or greater than B carry less weight in the individual’s perception than their numbers would justify. It is as if the individual looks through a *social filter* which exaggerates the importance of some and belittles the importance of others.
Let the density of the objective distribution be given by \( g(y) \) and the density of the perceived analogue by \( u_i(y) \) then the social filter is described by the expression

\[
\phi_i(y) = \frac{u_i(y)}{g(y)}
\]

If \( \phi_i > 1 \) for income greater than \( A \), the individuals with incomes greater than \( A \) get a disproportionately large reference weight, while individuals with incomes for which \( \phi_i(y) < 1 \) are underweighted. The social filter itself is sketched in figure 9.

![Figure 9: The social filter](image)

It looks on the log-income axis like a normal density with the top at \( \mu_i^* \) and dispersion \( \frac{\sigma_i}{\sigma_0} \) where \( \sigma_i \) stands for the welfare sensitivity of \( i \) and \( \sigma_0 \) for the log-standard deviation of the income distribution. We call the income class \( e^{\mu_i^*} \) the social focal point of the individual and \( q_i^2 = \frac{\sigma_i^2}{\sigma_0^2} \) the social myopia factor. On the basis of a modest sample (\( N=448 \)) created by Steven Dubnoff in the Boston area in 1983, Van der Sar (1988) estimated the model. It appeared that individuals have different social filters, which depend on a number of individual characteristics. He included as explanatory variables net family income, family size, the years of schooling of the respondent, and potential years of labor market experience. He found (pg. 91):

1. "Social myopia on average was about 2/3."
2. The myopia factor varies positively with schooling which reflects the fact that better educated people have more fantasy concerning income than the less educated and a broader social horizon, i.e., a broader reference group.

3. The myopia factor varies negatively with labor market experience.

4. The social focal point $\mu_i^*$ varies positively with own income which reflects the fact that the mode of the social filter function shifts with shifting income.

5. $\mu_i^*$ is positively correlated with family size indicating that the larger $f_s$, the higher the income needed.

6. Generally $\mu_i^* > \ln y_i$ which shows that people tend to focus their attention especially on people earning more than themselves; it may be explained by the fact that people are trying ‘to keep up with the Jones’s’; their mind is put on people being somewhat more fortunate; this view stresses the relative aspect.

7. The social filter is asymmetric in the sense that $B$ may give more reference weight to $A$ than vice versa.”

Then, in consideration of Van der Sar’s empirical results we can make the following propositions on the individual’s social reference group with respect to income:

I It especially contains those people with an income being (somewhat) bigger than one’s own.

II The width of the group is positively correlated with one’s years of schooling and negatively with one’s potential labor market experience; the effect of the latter being smallest in absolute terms.

This type of research has not been followed up yet, but there are now enough data sets to repeat and refine this at a larger scale.

The social filter also yields a definition of a social distance function on the log-income axis. More precisely we define the social distance between $i$ and $j$ as perceived by $i$ as $d_{ij}$ as

$$d_{ij} = \frac{1}{2\sigma_0^2} (\ln y_j - \mu_i^*)^2$$
which is essentially an expression derived from the social weight of
an individual: the higher the social weight \(=\phi_i(y_j)\) the lower the social
distance, that is, the less individual \(i\) is influenced by \(j\).

The first interesting aspect of the social distance is that it is asym­
metric: \(d_{ij}\) is generally not equal to \(d_{ji}\). As \(d_{ji}\) denotes the social distance
between \(i\) and \(j\) as perceived by \(j\), it can be seen as the passive distance
between \(i\) and \(j\). The active distance between \(i\) and \(j\) \((d_{ij})\) is a measure
of the influence which \(j\) has on \(i\), while the passive distance \(d_{ji}\) reflects
the influence of \(i\) on \(j\). Van Praag and Van der Sar (1991) calculated the
average social distance within and between schooling groups in the Boston
area:

<table>
<thead>
<tr>
<th>passive \ active</th>
<th>schooling 12</th>
<th>12&lt;schooling 16</th>
<th>schooling&gt;16</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>schooling 12</td>
<td>3.24</td>
<td>2.63</td>
<td>2.62</td>
<td>2.83</td>
</tr>
<tr>
<td>12&lt;schooling 16</td>
<td>4.00</td>
<td>2.27</td>
<td>1.98</td>
<td>2.78</td>
</tr>
<tr>
<td>schooling&gt;16</td>
<td>4.59</td>
<td>2.35</td>
<td>1.89</td>
<td>2.99</td>
</tr>
<tr>
<td>average</td>
<td>3.88</td>
<td>2.41</td>
<td>2.17</td>
<td>2.84</td>
</tr>
</tbody>
</table>

Source: Van Praag and Van der Sar (1991)

First we look at the diagonal elements. They present the within-
distance. We see that the average within-distance is the highest in the
group with low education, viz. 3.24, and the lowest in the group with high
education, viz. 1.89. The average distance between all subjects in the
sample is 2.84.

The passive distance of the lowest group to the middle-group, say
schooling inbetween 12 and 16 years of education, is 4.00 and the active
distance is 2.63. It follows that the low group exerts more influence on the
middle school group than that it is influenced by the middle-school group.

The social distance analysis is an instrument to study the segmentation
between social subgroups in society. In Van Praag and Van der Sar this
kind of analysis is not only performed on education classes but also on
income classes and age classes.

1.6.1 Discussion

The idea of the social filter model itself is more or less a truism. We
observe a norm on what constitutes a ‘good’, ..., ‘bad’ income level. If we
assume that such a norm is purely relative, the individual norm reflects a
subjective distribution of the norm’s subject, here the income level.
This idea has been put forward in sociology and psychology, e.g., by Runciman (1966), Easterlin (1974), Layard (1980), Kapteyn (1977), and Frank (1985).

The significance of this approach is in its application and the investigation into how the social filter function varies with individual characteristics and characteristics of the social environment.

Obviously, this kind of research is in its infancy. One of the obvious gaps in this presentation is that the social space $\Omega$ is here taken to be one-dimensional. The individual classifies other individuals by their income only. It is obvious that persons classify each other on a multiplicity of dimensions, e.g., age, education, type of employment, and religion. This would require a multi-dimensional social filter. A theoretical start has been made in Van Praag (1981) but due to a lack of data and computational manpower it has not been empirically operationalised up to now in a satisfactory manner.

1.7 Methodological discussion

The approach outlined above for measuring individual norms on income has been expanded to other aspects such as age and education by Van Praag, Dubnoff and Van der Sar (1988). More specifically, individuals were asked to connect age levels to subjective labels in the following Age Evaluation Question:

"When I think of other adults, I consider people to be

young, if they are younger than............. years old

somewhat young, if they are about......... years old

middle aged, if they are about............... years old

somewhat old, if they are about............. years old

old, if they are older than.................... years old"

Similar to the analysis of the IEQ, it is possible to analyse the age norms of respondents, for example, by explaining the answers by means of regression analysis. In Van Praag et al. (1988) this is done level by level
for the Boston data set. Let \( a_i \) (\( i=1..5 \)) be the respondent's age levels, then they consider the equation

\[
\ln a_i = \alpha_{0,i} + \alpha_{1,i} \ln \text{age} + \alpha_{2,i} \ln \text{schooling} + \alpha_{3,i} \ln f_s + \alpha_{4,i} D_{\text{gender}}
\]

where they assume that what is considered 'young' or 'old' depends on the age of the respondent, the number of years of schooling, the size of the family and the gender of the respondent. The results of the regressions are presented in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>constant</th>
<th>age</th>
<th>education</th>
<th>family size</th>
<th>gender</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>young</td>
<td>1.414</td>
<td>0.319</td>
<td>0.180</td>
<td>0.069</td>
<td>0.027</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>(0.270)</td>
<td>(0.043)</td>
<td>(0.067)</td>
<td>(0.026)</td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>somewhat young</td>
<td>2.329</td>
<td>0.266</td>
<td>0.045</td>
<td>0.056</td>
<td>0.019</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>(0.183)</td>
<td>(0.029)</td>
<td>(0.045)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>middle-aged</td>
<td>3.160</td>
<td>0.177</td>
<td>0.014</td>
<td>0.016</td>
<td>0.048</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td>(0.115)</td>
<td>(0.018)</td>
<td>(0.028)</td>
<td>(0.011)</td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>somewhat old</td>
<td>3.740</td>
<td>0.117</td>
<td>0.018</td>
<td>0.003</td>
<td>0.047</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td>(0.015)</td>
<td>(0.023)</td>
<td>(0.009)</td>
<td>(0.011)</td>
<td></td>
</tr>
<tr>
<td>old</td>
<td>4.243</td>
<td>0.058</td>
<td>0.067</td>
<td>0.003</td>
<td>0.048</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.099)</td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td></td>
</tr>
</tbody>
</table>

standard deviations in parentheses
Source: Van Praag et al. (1988)

From a statistical point of view, most coefficients are significant and follow a definite pattern. Our evaluation is that there is a strong systematic pattern which indicates that there is no confusion in connotation. The fraction of the variance explained, as measured by \( R^2 \), is poor in comparison to that of the IEQ, but is certainly not below standard for samples of microdata of this size (\( \approx 500 \)). However, it implies that there are more individual factors which were not covered in the survey, which must be added to the systematic structure than in the case of income standards.

With respect to the interpretation of the coefficients, we make the following observations. The older the respondent is, the higher his age standards. It follows that if A is 10% older than B, he will have an age standard for 'younger' that is about 3% higher (0.319*10%). Or in other words, if someone aged 20 finds himself 'somewhat young', an older person will find him still "young". For the age standards of "old", there is much less divergence between respondents of different ages.

We see that schooling has a strong impact on the definition of 'young': more educated people tend to stay 'young' longer. The impact of a large family on age standards is also evident. In such families, youngsters are considered to be children longer.
The social implications of these tendencies are not imminent. However, culturally it might be of interest that, in western countries, where the level of education has been increasing for decades, the concept of adulthood has become identified with an increase in age. Finally, the gender of the respondent also plays a role. If the respondent is female, the age standards are somewhat higher than for males, which implies that females tend to diminish the impact of age slightly. Except for the female tendency to stay and look young as long as possible, it is in conformity with the longevity of women compared to men.

Again we see that 'young' does not mean the same thing to young people as to old people. We can derive a general age \( a_i^* \) standard by setting \( a_i^* = \text{age} \), yielding

\[
\ln a_i^* = \frac{1}{1 - \alpha_{1,i}} [\alpha_{0,i} + \alpha_{2,i} \ln \text{schooling} + \alpha_{3,i} \ln f s + \alpha_{4,i} D_{\text{gender}}]
\]

The resulting age standards are tabulated in Table 5.

<table>
<thead>
<tr>
<th>Table 5: General age standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>General standards</td>
</tr>
<tr>
<td>male respondents</td>
</tr>
<tr>
<td>young</td>
</tr>
<tr>
<td>somewhat young</td>
</tr>
<tr>
<td>middle-aged</td>
</tr>
<tr>
<td>somewhat old</td>
</tr>
<tr>
<td>old</td>
</tr>
</tbody>
</table>

Source: Van Praag et al. (1988)

Similar to the age evaluation question, individuals were asked which education level they thought was 'very educated', 'uneducated', etc. By explaining the answers to this Education Evaluation Question (EEQ), Van der Sar (1991) was able to measure an individual norm on education as well. The interested reader is referred to Van der Sar (1991) for a full discussion of the AEQ, EEQ, and related questions.

### 1.7.1 Individual norms and general standards

The evidence described above suggests that people have subjective norms concerning various concepts. These norms will differ among individuals. They are measured by questions such as the IEQ, AEQ and EEQ, which supply us with numerical levels related to verbal labels or other symbols.
These questions may be posed theoretically in two ways: one may supply the label as stimulus and ask for an amount as a reply. Alternatively, one may supply an amount as stimulus and ask a label as a reply. The first way has been selected as the most practical when there are many different respondents with differing norms. It is also somewhat more informative, as people can space their answers.

In addition, we have evaluations by individuals of their own situation. This is done by fitting their own situation on their own norm. For instance an individual i with current income $y_c$ evaluates his own income by $U_i(y_c; y_c, f_s)$.

A final point is whether we may in some sense speak of general or social objective standards in contrast to individual subjective norms. Each individual may have an idea about what he thinks is a ‘bad’ or a ‘good’ income, but is there also a way to give content to a social norm with respect to what is a ‘good’ income and what is a ‘bad’ income? This question is especially pertinent when we think of a socially acceptable definition of poverty, or eligibility for social assistance. A second example is the general standard for being ‘old’, which is relevant for fixing the retirement age. A general standard may be derived from the individual standards by calculating the income level, age, education, etc., where people evaluate their own income, age, etc., as ‘bad’, ‘good’ etc. or ‘young’ and ‘old’.


In this method, there is a strong anchor effect. The answer of the respondent depends very strongly on his own situation. One may attempt to avoid this, for example, by asking

"Thinking about an average family with two children, what does it need per month for an adequate living?" (cf. Rainwater (1971))

While avoiding the anchor effect of one’s own situation, it introduces a new problem: what should be regarded as an average family, which will depend on the reference weighting system of the respondent. We can at
least deal with the anchor effect of the individual's situation, as we know the own situation of the respondent, but we do not know what the respondent considers to be an average family. The usefulness of this question thus depends on whether it is reasonable to assume that there is common agreement about what constitutes an average family. In heterogeneous populations, such agreement will be absent.

Obviously, the method works only to evaluate one-dimensional situations where numbers may be assigned and where a natural ordering is manifest.

A problem where the described IEQ method breaks down is when the society is only partly monetized. In that case, welfare cannot be characterized on the one-dimensional income scale. An ingenious way out has been suggested by Pradhan and Ravallion (1998). Their approach is to ask for evaluations of consumption levels instead of evaluations of income levels.

At present, welfare functions have been measured in nearly all EC countries, the USA, Hungary, Slovenia, Poland and Russia. In almost all cases, except in the USA, the samples were fairly large scale, ranging from a 1000 respondents to over 20,000. Panel data are scarce; the Dutch Socio-Economic Panel carried the question for a number of years, the German Socio-Economic Panel included the question in 1992 only, while at present, a Russian large-scale household panel includes the question as well. See the subsequent chapters for analyses of the latter two data sets.

1.8 Future directions: well-being and welfare

Traditionally, economists identify welfare (or even happiness) with income. However, it is well known and also fully recognised by other disciplines that there is more between heaven and earth than income and everything that can be bought with income.

This calls for an operational distinction between economic welfare and well-being. Welfare is the evaluation assigned by the individual to income or, more generally, to the contribution to our well-being from those goods and services that we can buy with money.

Next to material resources, we have other aspects which determine the quality of our life. We can think of our health, the relationship with our partner and family and friends, the quality of our work (job satisfaction), our political freedom, our physical environment, etc. We shall call this
comprehensive concept *well-being* or quality of life (see Nussbaum and Sen (eds., 1992) for philosophical discussions about this concept).

It is empirically possible for most individuals to evaluate their life as a whole. A well-known example is the following question devised by Cantril (1965):

> "Here is a ladder with ten steps which denotes the 'ladder of life'. The bottom step stands for the worst possible life. If you climb up and arrive at the tenth step, you arrive at the best possible life. Can you indicate where you are at the moment?"

Other questions which are very similar to Cantril's question ask individuals to denote how *satisfied* or how *happy* they are with their life as a whole. The concept of well-being is thus very similar to that of life-satisfaction or happiness, and we will not discuss the differences.

These questions are a standard module in many psycho-sociological surveys and respondents have no difficulty responding. See Veenhoven (1996) and Diener and Suh (1997) for reviews of the psychological literature on well-being. It is also obvious that responding to these questions is tantamount to evaluating one's life situation on a bounded numerical scale between zero and ten.

In fact, we have here a measurement method which *defines* the well-being concept in an operational way. We notice that what we measure is an evaluation of the individual's actual situation. Hence it is not an 'individual norm', as measured by the IEQ, where six qualitative labels are linked to income levels, yielding an 'income norm'. The Cantril question provides us with a *social standard* on well-being.

We assume again, as is always done implicitly, that the respondent's answers are comparable, in the sense that individuals evaluating their life with the same grade, such as a 5 or an 8, are equally unhappy or happy with their life. The main questions are:

a. What determines well-being?

b. What are the differences between welfare and well-being?

Plug and Van Praag (1995), Plug (1997), analysed these two questions on a large sample (1991) of about 6000 Dutch married couples, with the husband younger than 65 years of age. They related well-being to objective factors, like family size, income, age, and religion, and to subjective factors which relate to the intensity with which an individual 'has problems' with his health, job, marriage, physical environment, etc. Van Praag and Plug
(1995) then used the estimate to look at the welfare value of children and the well-being value of children. In general, the operationalization of well-being was found to be meaningful and yielded plausible results. We refer to Plug (1997) for a fuller discussion of the results.

In chapter 5, an attempt is made to explain well-being. In Chapter 6, the interpretation of well-being as a stream of momentary satisfactions is explored.

1.9 Conclusions

The work originating from the Leyden School has tried to operationalize the concepts of welfare, well-being, etc., which are considered immeasurable and esoteric by most of the economic profession. With rather simple and inexpensive questions in large-scale surveys, considerable information has been found on feelings. At least the feelings of welfare and well-being may be 'explained' by objectively measurable variables and by partial satisfaction measures with respect to aspects of life. The information is helpful for quantifying memory and anticipation weights and can be used to empirically identify social reference groups. The potential policy applications are plentiful. We briefly described its use to calculate family equivalence scales. A rather recent development is the combination of the welfare and well-being measurement, which makes it possible to identify the cost and benefits of various choices.

The apparatus developed thus far is not typically restricted to economic problems, but can also be used by psychologists, sociologists, and political scientists. Its use in health economics seems straightforward.

The story is hopefully not finished but only in the early stages. The main empirical restriction is that the data sets are scattered and almost never contain the IEQ, sound economic information (consumption, income, job characteristics) and at the same time 'soft' information on feelings on several aspects of life, such as the Cantril question. In this respect, the USA, where so much effort is given to research, is conspicuously absent.
Chapter 2

1.0 Conclusions

The work undertaken to examine this particular research question has involved the examination of a relatively small sample of 688 Dutch married couples. The sample is representative of the total population of married couples in the Netherlands. The research was conducted in 1960 and was based on a postal survey of married couples. The data was collected by means of a standardized questionnaire that was sent to each respondent. The questionnaire included questions on a wide range of topics, including the respondents' marital satisfaction, their perceived level of well-being, and their social contacts. The data was analyzed using statistical methods, including regression analysis and factor analysis. The results of the analysis showed that there were significant differences in the perceived levels of well-being between the respondents who were satisfied with their marriages and those who were not. The results also showed that there were significant differences in the perceived levels of well-being between the respondents who were satisfied with their social contacts and those who were not.

What are the differences between stress and well-being?

Plug and Van Prinse (1965), Plug (1967), analysed these two questions on a large sample (1961) of about 688 Dutch married couples, with the husband younger than 60 years of age. They divided well-being into objective factors, like family size, income, and location, and subjective factors which relate to the intensity with which an individual faces problems with his health, job, marriage, physical environment, etc. Van Prinse and Plug